Accident Analysis and Prevention 206 (2024) 107714

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Effect of spatial relationship between curves on crash severity at horizontal curves in a mountainous terrain

Deepak Awasthi *, Raman Parti, Kirti Mahajan

Department of Civil Engineering, National Institute of Technology, Hamirpur 177005, India

ARTICLE INFO

Keywords:
Adjacent curves
Spatial relationship
Curve attributes
Driver expectancy
Clustering
Statistical crash models

ABSTRACT

The study aimed to model the effect of spatial relationship between adjacent curves on the severity of curve-based crashes along with driver and crash causal characteristics, reflecting driver's short-term expectancy. The crash and other associated data was retrieved from the web-based Road Accident Data Management System available in Himachal Pradesh, India, and curvature attributes were extracted using GIS. Overall, the study included 1113 curve based crashes. The driver's perception of the sharpness of a curve was quantified as a single representative categorical factor based simultaneously on its length and radius using K-Medoid Clustering. Separate crash severity models catering to the two possible approach directions of the subject curve were developed reflecting its independent interaction with its corresponding adjacent curve in each direction. Partial Proportional Odds models were developed to overcome the predictive limitations of Ordinal and Multinomial logit models. Indicators of spatial relationship and the intensity of sharpness of the subject curve were found to be statistically significant. A sharp approach curve (radius:40–60 m) increased the risk of fatality by 2.16 times with a similar increase (2.5 times) observed for a short (length:30–60 m) adjacent curve. Adjacent curves turning in the same direction were 2.34 times more prone to fatalities. A very sharp subject curve with radius $\leq$ 40 m increased the risk of fatal crashes by 2.5 times, as did the short subject curves (30–60 m) (at least 3 times). Subject curves characterized by a short length and a very sharp curvature contributed relatively 3–4 times more to fatal crashes. The identified risk factors and their impact can help the relevant stakeholders to take appropriate actions and can further assist them in identifying high risk scenarios.

1. Introduction

Middle and low-income countries account for 93 % of global road fatalities despite only 60 % of the global traffic (Chen et al., 2020). The annual report on 'Road Accidents in India' for the year 2022 (Ministry of Road Transport and Highways (MoRTH), 2023) reported a total of 461,312 crashes resulting in 168,491 deaths and 443,366 injuries with rural areas accounting for 59.7 % of the crashes (NCRB, 2023). Mountainous highway construction is subjected to restrictions posed by topography, geological risk zones and concerns of environmental damage consequently leading to faulty and inconsistent geometric designs causing serious crashes (Li et al., 2023). Most studies in the past have found horizontal curves, a prevalent feature of rural roads in mountainous terrain, to be crucial from a safety standpoint. Fatalities on horizontal curves constituted 25 % and 28 % of the total crashes respectively in the United states (Gooch et al., 2016) and Thailand (Kronprasert et al., 2021) with Head-on and Run-off the road (ROR) being the most typical crash types (Torbic et al., 2004). For a driver, a combination of mountainous terrain and horizontal curves intensify the complexity further increasing the risk of severe crashes (Alrejjal et al., 2022; Chen et al., 2023; Meng, 2017).

The mountainous Himalayan region of India predominant in the northern and north-eastern states spans across 2500 km (Sustainable Development in The Indian Himalayan Region). These states collectively constitute 16.31 % of the total National Highway (NH) network, that provides primary interstate connectivity, with an average road density (road length in km/1000 square km) equal to 1.42 times the national average of 40.2. (Ministry of Road Transport and Highways (MORT&H), 2019). At collective mean death rate (deaths/1000 crashes) of 37.9, mountainous national highways have a poor safety record compared to the national average of 36.5. National highways documented 54,593 curve crashes in 2022, resulting in 20,573 fatalities and 55,866 injuries, a 10 %, 8 %, and 14 % rise respectively over 2021(Ministry of Road Transport and Highways (MoRTH), 2023). Thus, understanding the

* Corresponding author at: Department of Civil Engineering, National Institute of Technology, Hamirpur 177005, India.
E-mail addresses: deepak14awasthi@gmail.com (D. Awasthi), ramanp@nith.ac.in (R. Parti).

https://doi.org/10.1016/j.aap.2024.107714
Received 3 March 2024; Received in revised form 13 June 2024; Accepted 8 July 2024
Available online 14 July 2024
0001-4575/

contributing factors is essential for reduction of severity of curve based crashes.

A curve's interaction with other curves based on its relative location, generally referred to as their spatial relationship (distance from curve, radius of adjacent curve etc.), has been argued to influence its safety. In mountainous routes with frequent curves, spatial relationship can incorporate a driver's expectation of a future segment based on previous segments into safety assessments, unlike isolated curves (Findley et al., 2012). The higher likelihood of crash risk for some road segments highlights the importance of geometric design aspects in road safety with past studies relying on geometric design consistency to quantify the degree of agreement between driver expectation and road design (García et al., 2013). Spatial relationships have also been used to build geometric design consistency measures serving as a proxy to capture drivers' response to continuous changes in road geometry (Al-Sahili et al., 2018; Al-Sahili and Dwaikat, 2019; Jacob and Anjaneyulu, 2013). Moreover, different regions have their own unique road safety challenges due to unique conditions and driver behaviors, addressing which can enhance the overall understanding of road safety. This study contributes to the studies incorporating spatial relationships (or design consistency) along with crash and driver characteristics.

2. Literature review.

This section reviews past studies on design consistency assessment based on spatial relationships. Driver expectation and its role in affecting spatial relationships is then discussed followed by a discussion on crash modelling approaches. Then the research motivation and potential limitations of past studies are discussed.

2.1. Indicators of spatial relationship between curves

In the past, geometric design consistency has primarily been assessed using these approaches: (1) Individual characteristics of the adjacent segments, such as length, curvature, and distance to adjacent curves, (2) Operating speed consistency (3) Alignment indices and (4) Driving dynamics indices. Montella et al., (2008) considered a tangent distance greater than 250 m to interrupt the curve to curve sequence essentially rendering successive curves independent and concluded presence of a curve within 250 m to decrease the risk of severe crashes. Findley et al., (2012) examined the spatial relationship using adjacent curve's features, distance, and turn direction, however only distance was employed to develop crash prediction models (CPMs). The observed and predicted collision rates were more closely aligned with CPMs based on spatial relationships. Gooch et al., (2018, 2016) found a substantial reduction in crash risk by an increase in the degree of curvature of the adjacent curve within 0.75 miles (1207 m) which was however insignificant beyond 0.75 miles. Elvik, (2019) found reduction in distance between adjacent curves to decrease the crash rate on relatively sharp curves (radius:100 m).

In addition to the operating speed consistency (Chaudhari et al., 2022; Jacob and Anjaneyulu, 2013), recent studies have employed Inertial Consistency (ICI) to quantify crash risk. ICI refers to the difference in the operating speed at a curve's entry and operating speed based on preceding segments averaged over a finite distance (García et al., 2013; Llopis-Castelló et al., 2018a; Montella and Imbriani, 2015) or time ((Llopis-Castelló et al., 2018b, 2018c), refer Table 1, with a higher ICI corresponding to a lower design consistency (DC) and a higher crash rate.

Alignment consistency indices (AC), particularly the global alignment indices based on an entire highway segment, have been extensively employed for safety assessment (Al-Sahili et al., 2018; Al-Sahili and Dwaikat, 2019). Khan employed 17 design consistency metrics, including AC, to assess fatal and injury collision risk and found them statistically significant, concluding AC based measures as more essential than individual curve features. Further it was concluded that

Table 1
Indicators of spatial relationships and safety at curves in the past studies.

Indicators of Spatial Relationship	Reference	Analysis Approach	Inference/ Remarks
(1). Individual Characteristics of adjacent curves or their hybrid combination			
Presence of Adjacent Curve	Montella et al., (2008)	CPMs: NB based models	Presence of a curve within 250 m (as a binary variable) significantly decreased the risk of severe crashes.
Distance to Adjacent Curve	Findley et al. (2012), Khan et al. (2013)	Quasi Poisson and Negative binomial (NB) based CPMs.	Reduction in curve separation reduced the overall crash risk indicating relatively closer curves to be safer
Radius or Degree of curvature of Adjacent Curve	Findley et al. (2012)	CPMs: NB based models	Not included in the final model
Length of adjacent Curve	Findley et al. (2012), Gooch et al. (2018, 2016)	CPMs: NB based models	Insignificant or uncorrelated with crash frequency.
Relative Direction of turn of Adjacent Curve	Findley et al. (2012)	Correlation analysis	Uncorrelated with crash frequency.
Degree of Curvature of adjacent curve as a function of its distance from study curve	Elvik (2019), Gooch et al. (2018, 2016)	CPMs: NB Models, Relative Risk Ratio	Crash Risk reduced with increase in curvature of proximal curves with the effect being insignificant for curves that were not proximal
(2). Operating Speed Consistency			
Distance based ICI	García et al. (2013), Montella and Imbriani (2015), Llopis-Castelló et al. (2018a)	ICI developed based on preceding segment length of 1000 m, 5000 m and 600 m, CPMs: NB based Models	Developed ICI measures showed good performance in estimating crash risk
Time based ICI	Llopis-Castelló et al. (2018b), Llopis-Castelló et al. (2018c)	ICI based on time (t) = 15 s, CPMs: NB based Models	Time based ICI outperformed distance-based ICI in crash risk estimation.
(3). Alignment Indices			
Global Alignment Indices: $L_c = L_{curve(T)}/L_{Section}$, R_{Max}/R_{Min}, CRR $= R_{curve}/R_{Avg}$, R_{Avg}, Relative Area (R_a), Operating Speed dispersion, $R_D = R_{Avg}/R_{VD}$	Polus and Dagan (1987), Polus and Mattar-Habib (2004), Al-Sahili et al. (2018), Al-Sahili and Dwaikat (2019), Azmeri Khan et al. (2023), Khan et al., (2023)	Crash rate: Linear Regression, CPMs: NB based models	Higher inconsistency increased crash risk.
Local Alignment Indices: 1. RTR = Avg $[T_{US},T_{DS}]/R_{curve}$ 2. CRR_M = Avg $[R_{US},R_{DS}]/R_{curve}$	Hamilton et al. (2019)	CPMs: NB based Models	Proximal sharp curves reduce overall crash risk on subject curve. Crash risk reduced with reduction in the separation of adjacent curves.

*Notations used:
Lc = Relative Length of Curves in a section.
Lcurve(T) = Length of curves in the section.

LSection = Length of the section.
RMax = Maximum radius of the section.
RMin = Minimum radius of the section.
ICI = Inertial Consistency Index.
Rcurve = Radius of the study curve.
RAvg = Average radius of the section.
RVD = Minimum radius of the section based on design speed.
Avg = Average.
TUS = Upstream tangent.
TDS = Downstream Tangent.
RUS = Radius of the Upstream Curve.
RDS = Radius of the Downstream curve.
CRR_M = Modified Change Radius Rate.

consistency between successive segments influences safety. Combining roadside hazard and similar design consistency measures, Azmeri Khan et al. (2023) analyzed ROR crash severity but reported higher variations in curve radii in the overall segment to reduce the crash severity. Compound curves, reverse curves, left turning curves, longer tangents between curves, and shorter intra-curve distance between compound and reverse curves increased the risk of fatal single vehicle ROR crashes (SV-ROR), according to Chakraborty and Gates, (2023). Hamilton et al., (2019) developed alignment indices accounting for the influence of the modified change radius rate (CRR_M) and average tangent distance of nearby curves on both sides relative to subject curve radius (study curve/or curve of interest) (RTR), to quantify crash risk. Increasing CRR_M significantly increased crash risk on the subject curve, indicating flatter adjacent curves to improve curve safety. However, the analysis did not account for adjacent curves closer than 90 m (300feet) (Hamilton et al., 2019). Table 1 summarizes the various indicators of spatial relationship and their relationship with safety.

2.2. Driver expectancy

Expectations based on previous segments affects driver's safety perception and any inconsistency violating it can result in risky scenarios. Operating speed-based consistency models have been used for global and local consistency evaluation. Global consistency models based on the speed variation at the overall alignment/segment level have been used for developing CPMs. Local consistency concerns itself with sudden transitions between successive elements (tangent to curve transitions or curve to curve transitions) and prioritizes safety treatments by identifying high risk locations. García et al., (2013) and Montella and Imbriani, (2015) developed ICI using a continuous operating speed profile respectively for the past 1000 m and 5000 m driven accounting for long-term driver expectation. However, adults are expected to lose the previously acquired information within 18 s(s) based on "Short-Term Memory" (STM) (Revlin, 2012). Different distance and time intervals (distance: between100 and 1500 m, time: between 10–40 s) have been used to develop the corresponding distance and time-based ICI measures for design consistency evaluation to reflect STM-based short-term driver expectation (Llopis-Castelló et al., 2018a, 2018b, 2018c). Global consistency models based on distance-based ICI corresponding to 600 m produced the closest agreement with short term expectation (Llopis-Castelló et al., 2018a) but were outperformed by time-based ICI measures distance in capturing short term expectation specifically for time = 15 s (Llopis-Castelló et al., 2018b). However, operating speed-based design consistency models can differ based on vehicle type making it difficult to have a single representative design consistency measure. Thus, local consistency evaluation using adjacent/successive features (curve or tangent) and short-term expectation can accurately represent driver's safety perception.

2.3. Crash severity analysis techniques.

Broadly two modelling approaches form the core of crash safety analysis: Crash prediction models (CPMs) based on crash frequency and Crash severity modelling. CPMs (Negative binomial, Poisson based, Generalized linear models etc.) help investigate contributory factors whereas crash analysis further helps in designing targeted countermeasures (Ferreira-Vanegas et al., 2022; Jeong et al., 2018; Li et al., 2012; Santos et al., 2022; Savolainen et al., 2011; Wen et al., 2021). Previous studies have employed various statistical (Binary logit/probit (BP), Ordered logit/probit (OL/OP), Multinomial logit (MNL), Random/Mixed parameters etc.) and machine learning (ML) models (Support vector machine (SVM), Random Forest (RF), Neural Network (NN) etc.) to model crash injury severity (Ferreira-Vanegas et al., 2022; Jeong et al., 2018; Santos et al., 2022; Savolainen et al., 2011; Wen et al., 2021) with reasonable model performance. Despite better prediction accuracy of ML models compared to statistical models (Ahmadi et al., 2020; Iranitalab and Khattak, 2017; Li et al., 2012), their black-box nature makes them difficult to interpret. Whilst the MULTI models fully relax the proportional odds assumption (PO) and allow the effect of explanatory variables to vary independently across each severity level, they assume unordered severity levels. On the other hand, OP/OL models (Chen et al., 2023; Ma et al., 2022) account for ordered nature but are limited by PO assumption. Nested logit model (NL) permits the variable effects to vary through the injury levels, however the nested structure's interpretation and identification is complex (Azmeri Khan et al. (2023); Mooradian et al., 2013). Regardless, these traditional techniques consider constant parameters failing to account for unobserved heterogeneity (Song et al., 2023; Yu et al., 2019). Random parameter (or effect) models like Mixed logit models, (RP-Ordered Probit (RP-OB), RP-Hierarchical Ordered Probit (RP-HOP) etc. accommodate unobserved heterogeneity (Azmeri Khan et al. (2023)). A combination of decision tree and RP-HOP produced the best fit while simultaneously accounting for unobserved heterogeneity (Azmeri Khan et al. (2023)). However, traditional random parameter models assume independent random distribution for explanatory variables, ignoring their interaction effect on random parameter mean and variance. Additionally, temporal instability of factors has been a major source of unobserved heterogeneity and if remain unaccounted can affect the long-term effectiveness of safety measures. Yu et al., (2019) used Latent class mixed logit model with temporal indicators to assess temporal instability-induced heterogeneity and found it to accurately capture marginal effects based on time-based crash severity analysis. Song et al., (2023) employed an improved random parameter bivariate probit model with heterogeneity in means with temporal indicators to simultaneously handle unobserved heterogeneity and temporal variations. However random parameter-based models are very complex, non-transferable and observation specific (Lord and Mannering, 2010).

Partial proportional odds (PPO) model has also been used for crash severity analysis in the past (Ma et al., 2022; Meng, 2017). PPO model bridges the gap between ordered and unordered models by accounting for the ordered severity levels and partially relaxing the PO assumption for some variables making it practically more realistic. PPO models have been shown to outperform MULTI, OL and Mixed Logit models in terms of overall fit (Jou and Chao, 2022; Mooradian et al., 2013; Yu et al., 2022). PPO model was thus used in the present study as it produces stable predictions on unseen data with an excellent predictive performance especially with smaller sample size (Meng, 2017; Mooradian et al., 2013).

2.4. Research motivation

As per literature, despite significant effect of spatial relationships between curves on road safety, its role in influencing crash severity has not been sufficiently explored (refer Table 1). Local consistency in alignment based on spatial relationship between adjacent elements can accurately capture drivers' environmental response further revealing their safety perception. A variety of distances between consecutive curves ranging from as close as 0.26 miles (418 m) (Findley et al., 2012) to as far as 1.25 mile (2011 m) (Elvik 2019) have been used to define them as proximal. However, studies past rarely accounted for the influence of preceding curve on crash risk (count or severity) based on driver's short-term expectation and curve-to-curve sequence continuity, which is especially important for scenarios involving a very close-knit series of curves before a crash. Therefore, defining the proximity of curves through a limiting distance (or range of distance) between adjacent curves might be beneficial in such a scenario. Further, the overall intensity of sharpness experienced by a driver navigating a curve is a resultant of the combination of its length and radius. Moreover, instability of the vehicular forces at such close-knit curve sequence especially due to their relative direction of turn further adds to driver workload (Al-Sahili et al., 2018; Wang et al., 2013). To address these limitations, the current study incorporated the interaction between radius and length as the "Intensity of sharpness" for a curve and relative direction of turn between adjacent curves.

The study aims to model crash severity for a given crash on the subject curve. While traversing a close-knit series of curves, if the driver (at fault) crashed just before exiting the subject curve, there might be higher influence of subject curve whereas if the crash occurred at the subject curve's entry, the influence of adjacent/preceding curve would be considered stronger among the two curves. Moreover, driver behavior and expectancy are influenced by the adjacent curve based on the direction of approach. Hence if information of the approach direction is not available, separate models are needed to analyze the independent interaction between the subject curve and its adjacent curve in influencing the resultant severity, depending on the direction from which the subject curve was approached. Additionally, driver related factors such as driver demographics, erroneous behaviors and other crash causal factors can influence crash severity (Ahmadi et al., 2020; Jeong et al., 2018; Kim et al., 2022; Llopis-Castelló et al., 2018b, 2018c). Therefore, the present study attempts to contribute to the limited studies on curve based crashes by investigating the role of spatial relationship in terms of curve proximity incorporating driver expectancy, curve attributes and other crash causing factors on crash severity.

2.5. Aim and objectives

The present research aims to model the severity of curve-based crashes based on spatial relationship between adjacent curves simultaneously incorporating the driver and causal factors involved, given a crash has occurred. The objectives of the study are: (1) to study and compare the individual and the interactive effect of radius and length of a curve in influencing crash severity, and (2) to model the severity on the subject curve based on its independent spatial relationship with its corresponding adjacent approach curve in a given direction, traversed immediately prior to the crash along with the driver and crash characteristics.

3. Methodology

Crash information and geometric data were extracted followed by categorization of curve attributes along with characterization of the intensity of sharpness. Crash severity with three levels was then modelled based on the relevant factors. The impact of the resultant statistically significant factors on severity was then analyzed based on the predicted rate of the severity levels. The entire analysis was performed using R programming.

3.1. Study area and crash data collection

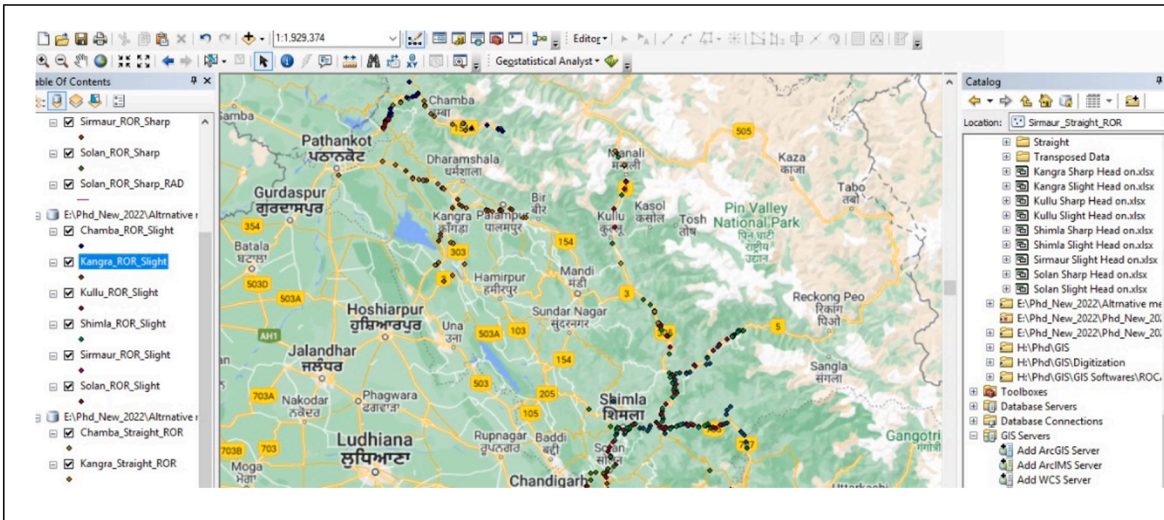
Two-Lane NH in mountainous rural areas (non-snow bound) of Himachal Pradesh, India, with an average death rate of 39.7 (Ministry of Road Transport and Highways (MoRTH), 2023) were chosen for the present study. In terms of crash count and deaths for 2022, the state ranked second and third respectively among the mountainous states with 2484 crashes resulting in 979 deaths (NCRB, 2023). The state's Road Accident Data Management System (RADMS) (introduced in 2015) was used to extract 1319 curve-based single vehicle ROR and Head-on crashes from 2016 to 2022, representing 97 % of the total curve-based crashes. The analysis includes crashes classified as "Fatal," "Grievous (i.e., incapacitated for longer periods)," and "Injury needing hospitalization" (or simply hospitalizations) (non-incapacitating) in decreasing order of severity. The damage only crashes were significantly underreported and were excluded from the final analysis (Tiwari et al., 2022; Tiwari and Mohan, 2016). The database contained precise geographical coordinates of these curves and additional crash and at-fault driver information. Due to the lack of road inventory data (typical in LMIC countries), the required curvature attributes (radius and length) were extracted from developed curve shapefiles using an ESRI ArcGIS based toolbox named 'Road Curvature Analyst' (Bíl et al., 2018; Nair and Bhavathrathan, 2022). Google street map-based platform was used to digitize the shapefiles. Fig. 1 shows the discretization and extraction of curvature attributes.

Based on geographic orientation and direction of approach, the 'upstream' (i.e., south or west of the subject curve) and 'downstream' (i.e., north or east of the subject curve) were defined in the study (Hamilton et al., 2019). The influence area of the preceding curve for defining it as adjacent or proximal was finalized as 210 m (+20 m tolerance) from the subject curve based on a travel time corresponding to a speed limit of 50kmph (refer Table 2) for 15 s (short term expectancy), additionally maintaining the curve-to-curve sequence. This yielded 822 subject curves (62.32 %) that had an adjacent curve within 170 m (94 % within 120 m) on both its sides and were termed "serial curves". Only 91 curves (6.90 %) had no adjacent curve on either side at a minimum tangent distance of 450 m rendering them completely independent and were termed "isolated". Isolated curves were removed from further analysis due to insufficient sample size to draw any reliable conclusion even if one uses the independent tangent length as a factor. Interestingly 406 curves termed as "leading curves" had an adjacent curve within 170 m (92 % within 120 m) on one side and an independent tangent greater than 250 m on the other. However, the distribution of the independent tangent was very narrow (250–290 m). The individual relationship with injury severity was insignificant despite several discretization attempts based on Chi-squared test (Chi-sq) for association. Moreover, after removing isolated curves, 100 % of the subject curves had an adjacent curve within 170 m, predominantly within 120 m. However, after removal of outliers, unrealistic radius and length combinations, characterization of the "Intensity of sharpness" (refer 3.2.2) and considering the independent relationship of subject curve with its corresponding preceding curve in a given direction, the final dataset consisted of 745 serial curves and 368 leading curves corresponding to an overall crash count of 1113 crashes. A serial subject curve would be shared by datasets in both upstream and downstream directions. This consequently resulted in 927 curve-based crashes (745 serial and 182 leading) in the upstream dataset with 931 subject curves (745 serial and 186 leading curves) in the downstream.

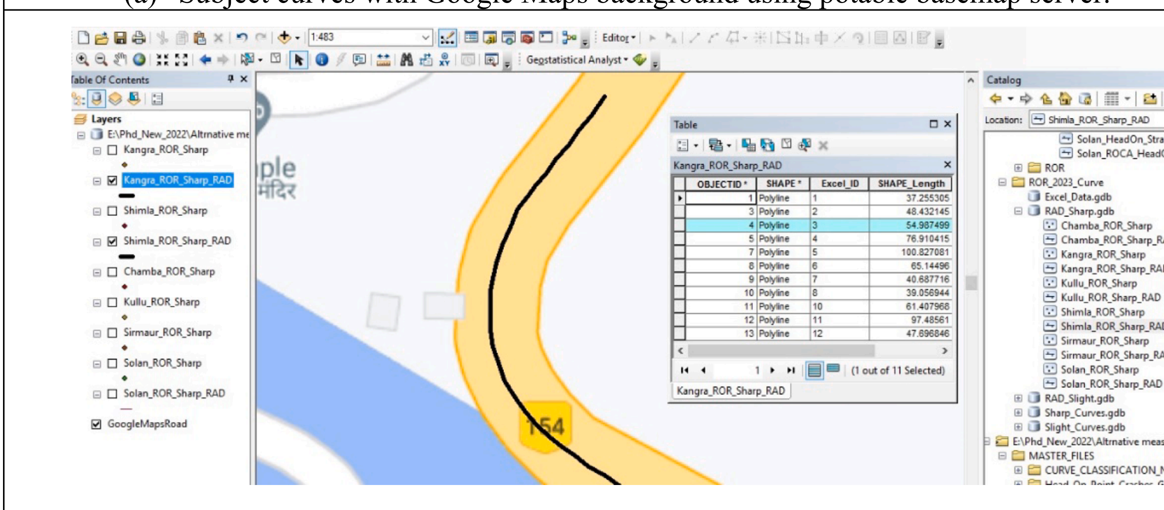
3.2. Data extraction.

3.2.1. Crash and driver related factors

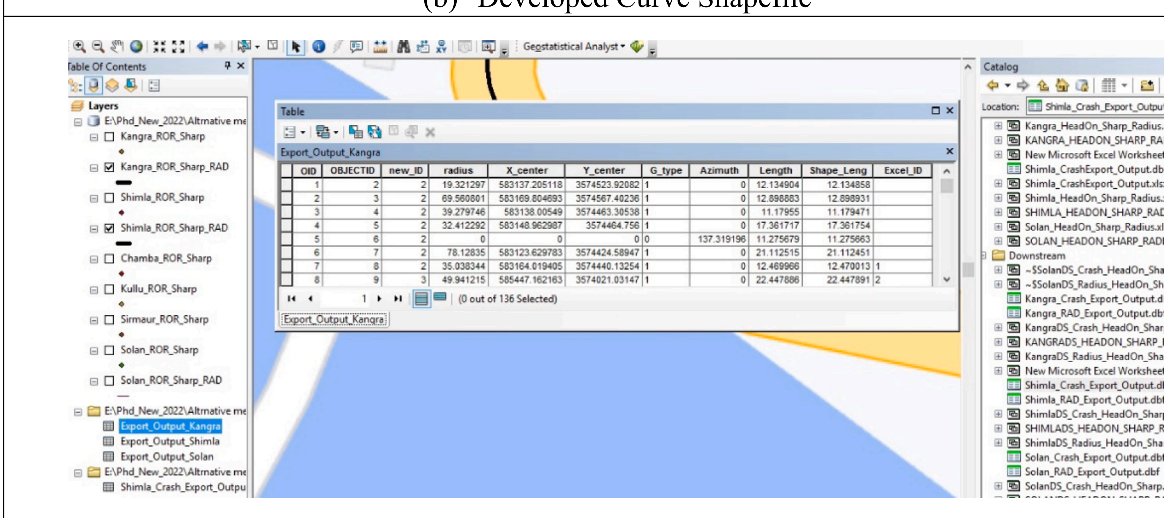
The crash, driver and temporal characteristics including at-site reported weather, day, lighting condition, vehicle type, speed limit, at-fault driver's age, and crash cause were extracted from RADMS and categorized as shown in Table 2. The at-site recorded crash causes were divided into five categories (Table 2). Dangerous driving included driving against the flow, ignoring right of way restrictions, and other risky maneuvers. For simplicity, crashes caused by malfunctioning/lack of safety equipment like parapets or guardrails, slippery road surface, inadequate markings, blind curves, etc. were replaced with 'Road infrastructure and alignment' throughout the article. A 'Careless Turn' occurred when the motorist failed to anticipate or lost control of the



(a) Subject curves with Google Maps background using potable basemap server.



(b) Developed Curve Shapefile



(c) Curvature Extraction from ROCA

Fig. 1. Curvature extraction from digitized curve shapefiles using ROCA.

Table 2
Description of Driver factors, Crash Related Factors, Radius and Curve Length Data.

(1). Driver and crash related factors (Categorical Factors)

Categorical Factor	Subcategory	Upstream Dataset		Downstream Dataset	
		Count	% Share	Count	% Share
Curve_Direction	Opposite	618	66.6	621	66.70
	Same	309	33.33	310	33.30
Day_of_Week	Weekday	610	65.80	642	68.96
	Weekend	317	34.20	289	31.04
Weather	Cloudy	58	6.26	55	5.91
	Fine	781	84.25	791	84.96
	Precipitation_or_Fog	88	9.49	85	9.13
Light_Condition	Dark	272	29.34	272	29.22
	Darkness_with_street_light	32	3.46	32	3.44
	Day_Light	564	60.84	572	61.44
	Twilight	59	6.36	55	5.90
Crash_Cause	Careless_turn	92	9.92	87	9.34
	Dangerous_driving	280	30.20	281	30.18
	Overtaking_or_Lane_change	133	14.35	142	15.25
	Road_Infrastructure_and_Alignment	82	8.85	80	8.59
	Speeding	340	36.68	341	36.64
Vehicle_Type	HMV (Heavy Motor Vehicle)	187	20.17	192	20.62
	LMV (Light Motor Vehicle)	578	62.35	580	62.30
	Two_Wheeler	162	17.48	159	17.08
Speed_Limit	≤30kmph	572	61.70	587	63.05
	30kmph-50kmph	251	11.22	237	11.49
	≥50kmph	104	27.08	107	25.46
Driver_Age	≤30 (Less_eq_30): (Young)	411	44.34	431	46.29
	30–50 (30_to_50): (Middle Aged)	140	15.00	96	10.32
	≥50 (Greater_eq_50): (Old)	376	40.56	404	43.39

(2). Curve attributes in the Upstream (US) and Downstream Datasets (DS)

	Upstream Dataset				Downstream Dataset			
Curve attribute	Mean	Std. Dev	Max	Min	Mean	Std. Dev	Max	Min
Radius of the Subject curve (RAD)	51.59	23.2	149.6	14.2	52.4	23.1	144.0	12.3
Length of the Subject Curve (LEN)	66.5	27.6	148.7	12	66.1	66.1	148.7	14
Approach curve Radius (US_RAD or DS_RAD)	58.0	25.9	145.7	16.4	55.8	55.8	149.9	11.5
Approach curve Length (US_LEN or DS_LEN)	64.6	25.5	148.8	148.8	61.3	61.3	148.3	16.8

curve. The radii and length were extracted using ROCA and the relative direction of turn of the subject and adjacent curve (Curve Direction) was categorized into two categories namely: same and opposite. The resulting dataset revealed that both the length and radius were invariably less than 150 m. Table 2 shows the descriptive statistics of the extracted crash data, driver data and curve attributes for the final dataset.

3.2.2. Categorization of curve attributes (radius and length) and characterization of intensity of sharpness based on K medoid clustering

Both the individual curve attributes and the intensity of curve's sharpness were separately characterized for subsequent use in the "No interaction" and "Interaction" models respectively (refer 3.3). Considering their narrow distribution, K medoid clustering, an unsupervised ML algorithm, was used to cluster natural subgroups in radius and length separately for upstream and downstream datasets. K medoid clustering, is resilient to outliers, but the number of clusters (k) needs be defined in advance (Ikotun et al., 2023; Vinod, 1969). The resulting output is a cluster medoid corresponding to the input number of clusters (k) in the algorithm. The optimal cluster count was determined based on the statistical significance of the relationship (if any) of the resulting clusters with injury severity based on a Chi-Square (chi-sq) test of association. Although there are no defined rules for categorizing curvature in India, data distribution features and IRC 52:2019 (Indian Roads Congress, 2019) were used for the cluster descriptions as shown in Table 3. A resultant cluster was named based on the range in which the cluster medoid belonged and were combined into a single cluster, in case of an overlap. Table 4 shows the selected optimal number of clusters and their description. The radius of the subject curve, its length, and the length of the approach curve in the upstream were found to be significantly associated with injury severity.

Table 3
Nomenclature Scheme for the Categorization of Length and Radius of a Curve.

Length (m)	Nomenclature	Radius (m)	Nomenclature
0–30	Tight Length Curve (TLC)	**0–40**	Very Sharp (VSS)
30–60	Short Length Curve (SLC)	**40–60**	Sharp (SHS)
60–90	Medium Length Curve (MLC)	**60–80**	Marked Sharpness (MKS)
>90	Long Length Curve (LLC)	**>80**	Significant Sharpness (SFS)

A curve's intensity of sharpness was characterized as a single representative categorical variable combining its length and radius based on K medoid clustering. The resultant output is a pair of radius and length medoids (length, radius). The optimal number of clusters were selected based on three criteria: Cluster Silhouette (Rousseeuw, 1987), Cluster size balance and significance of the chi-sq association of the resulting clusters with the injury severity. Whereas higher cluster silhouettes indicate better clustering and separation, imbalance in clusters due to excessive or insufficient sample sizes on the other hand indicate poor clustering performance. Table 4 shows the resultant optimal clusters for the two datasets for the intensity of sharpness of the subject curves and approach curves in the respective datasets and the medoids of the resulting clusters were described using Table 3.

3.3. Development of crash severity models

The driver and crash factors described in Table 1 were used to develop the models along with the curvature attributes (radius and length) of the subject curve (Table 4). Furthermore, the indicators of

Table 4
Categorization of Curve Attributes (radius and length) and Characterization of the Intensity of Sharpness.

(1) Categorization of the Individual Curve Attributes (radius and length) in the Upstream (US) and Downstream (DS) Datasets based on K Medoid Clustering.

(a). Individual Factors	Upstream Dataset				Downstream Dataset			
	Optimal no. ofClusters	p-value (chi-sq)	Cluster Medoids	Category	Optimal no. ofClusters	p-value (chi-sq)	Cluster Medoids	Category
Radius of the Subject Curve (RAD)	4	0.073	47.1	SHS	4	0.060	50.3	SHS
			71.8	MKS			27.6	VSS
			26.0	VSS			76.3	MKS
			95.5	SFS			100.9	SFS
Length of the Subject Curve (LEN)	3	0.000	72.6	MLC	3	0.000	77.8	MLC
			113.6	LLC			117.2	LLC
			39.73	SLC			41.3	SLC
Approach Curve Radius (US_RAD or DS_RAD)	4	0.208	67.3	MKS	4	0.648	32.3	VSS
			99.4	SFS			69.9	MKS
			30.4	VSS			49.1	SHS
			46.9	SHS			102.8	SFS
Approach curve Length (US_LEN or DS_LEN)	3	0.001	112.8	LLC	3	0.191	68.7	MLC
			74.2	MLC			43.0	SLC
			40.7	SLC			97.5	LLC

(2) Characterization of the Intensity of Sharpness for the Subject and Approach Curve in the Upstream (US) and Downstream (DS) Datasets based on K Medoid Clustering.

Category of the Curve	Optimal number of Clusters	Cluster size	Silhouette	p-value (chi-sq)	Cluster Medoids		Intensity of Sharpness
					Length	Radius	
(a) Intensity of Sharpness for the Subject Curve and Upstream Approach Curve (Upstream Dataset)							
Subject Curve (RAD_LEN)	4	284	0.372	0.001	72.8	45.9	MLC-SHS
		171			107.6	66.6	LLC-MKS
		299			43.1	31.3	SLC-VSS
		173			51.2	73.8	SLC-MKS
US Approach Curve (USRAD_USLEN)	5	134	0.343	0.046	102.1	64.4	LLC-MKS
		149			77.9	96.9	MLC-SFS
		250			39.0	36.7	SLC-VSS
		215			69.3	44.3	MLC-SHS
		179			50.5	63.4	SLC-MKS
(b) Intensity of Sharpness for the Subject Curve and Downstream Approach Curve (Downstream Dataset)							
Subject Curve (RAD_LEN)	4	280	0.353	0.002	72.1	43.8	MLC-SHS
		185			107.5	66.5	LLC-MKS
		255			40.7	30.1	SLC-VSS
		211			54.5	70.1	SLC-MKS
DS Approach Curve (DSRAD_DSLEN)	4	226	0.348	0.052	74.1	44.9	MLC-SHS
		243			56.1	70.8	SLC-MKS
		342			44.2	36.4	SLC-VSS
		120			95.4	91.6	LLC-SFS

spatial relationship for modeling included the adjacent curve's curvature characteristics and their relative turn direction. Separate models were developed for the two datasets corresponding to the two directions. For a given direction, two separate models were then further developed. The first model explored the spatial relationship based on the individual curvature attributes of the subject and its adjacent curve without combining the radius and length, termed as the "No interaction" model, along with other driver and crash factors. The individual curve characteristics were replaced with the intensity of sharpness in the second model, termed as the "Interaction model". The model's performance was evaluated based on its fit on the train set and unseen test set predictions. Each dataset was divided into two groups: 85 % for training and 15 % for evaluating (or validating) the prediction performance. Stratified sampling was used for both train and test sets to ensure proportionate representation of the three severity levels across the segmented train-test sets, similar to the unsegmented datasets, ensuring a more representative multiclass prediction evaluation. The respective proportional representation of the three severity levels i.e., fatal, grievous, injury needing hospitalization in the upstream and downstream dataset respectively were (29 %, 29 %, 42 %), and (28 %, 28 %, 43 %). Goodness of fit of the statistical models was assessed based on AICc (a more suitable version of AIC for smaller datasets), a Likelihood ratio test (LRT) and pseudo R^2. The LRT test statistic is chi-square distributed (Deng et al., 2006) where a significant value indicates a significant improvement over a null model. The model's average overall accuracy, sensitivity (i.e., capacity to correctly identify examples of a certain class), and specificity (i.e., ability to accurately reject instances that do not belong to that class) were used to evaluate its class-wise prediction capability in both the train and test sets. Instances were assigned severity level for which the estimated probability of belonging to that level was the highest amongst the three severity levels. An ordinal logit model was first trained which showed the violation of PO assumption. Consequently, PPO model was trained that outperformed MULTI model on comparison.

3.3.1. *Ordinal logit model (ORDI model)*

Ordinal logit model or ORDI model estimates the cumulative odds of an instance to belong in or below a particular category based on the predictor variables. The estimated coefficients depict how change in predictor variables (x) impact the odds of an outcome (y) to belong in or below a given category. The following equation represents the cumulative odds of an instance to fall in or below a particular category:

$$\mathrm{Log}\left(\frac{P(y \leq i)}{1 - P(y \leq i)}\right) = \mathrm{ai} - \mathrm{b}^{\mathrm{T}}\mathrm{X}\mathrm{where} \tag{1}$$

$P(y \leq i)$ is the cumulative probability of being in or below a particular category i, a_i is the intercept for category I and b represents a vector consisting of coefficients for predictor variable set X. T refers to matrix transpose. ORDI model was developed using "polr" function in "MASS" package in R. Table 7 shows the model performance metrics for the developed models. The probability for an instance to belong to each class/level of the dependent variable can be extracted from the output of 'polr' function.

3.3.2. Multinomial logit model (MULTI model)

A MULTI model used for multiclass classification, completely relaxes the PO assumption but does not account for the inherent order (if any) in the dependent variable. The probability of falling into a category j for a given crash is given by (Liu et al., 2021; Ye and Lord, 2014):

$$P(Y_i > j) = \exp(\beta_j{}^*x_i)/(\sum_{i=1}^{j}\exp(\beta_j{}^*x_i) \quad (2)$$

MNL models were developed using 5-fold cross validation (caret package). Table 7 shows the performance metrics of the developed MULTI models.

3.3.3. Partial proportional odds model (PPO model)

PPO model addresses the limitations associated with ORDI model by considering the ordinal nature of severity and relaxing the PO assumption for variables that reject the null hypothesis in the brant test, thereby allowing their effect to vary across different severity levels. The probability of injury severity level j for a crash event i can be estimated as (Liu et al., 2021):

$$P(Y_i > j) = \exp(\alpha_j + \beta_j x_i)/(1 + \exp(\alpha_j + \beta_j x_i)), \textit{where } j = 1, 2, 3 \ldots (j-1) \quad (3)$$

The "vglm" function in "VGAM" package in R was used for PPO model. A positive estimate increased the likelihood of remaining in or below a particular category compared to the reference category (Fatal in PPO model). To put simply, a positive estimate implies a likelihood to "remain in or move to a lower injury subcategory". Table 5 and 6 represent the estimates for the contributing factors in the upstream and downstream models respectively while Table 7 shows the performance metrics of the developed PPO models.

4. Results

4.1. Selection of the best model

An initial run of the ORDI models revealed Accident Day, Weather and Speed Limit to be consistently insignificant across both the datasets (upstream and downstream) and were hence dropped from the subsequent analysis. Despite performing reasonably well on the test and train sets, the ORDI model estimated a zero sensitivity and a unit specificity for grievous crashes indicating its failure to learn the underlying patterns (Table 7). A Brant test (Liu et al., 2021) revealed a significant violation of PO assumption. An in-depth investigation revealed vehicle type, radius of the subject curve (RAD) and driver age to significantly violate the PO assumption in the downstream ORDI models while vehicle type, upstream radius (US_RAD) and upstream length (US_LEN) did so in the upstream ORDI models, indicating their impact to likely differ across severity levels. PPO model relaxing the PO assumption for the above variables was trained and compared to MULTI model that relaxes the assumption for all variables.

On comparison, the PPO models outperformed the MULTI models in terms of AICc, implying a relatively better fit (Table 7). The MULTI models, although produced similar results for other factors in general, led to counterintuitive results across both the datasets for the radius of the subject curve in contrast to the PPO model that produced more intuitive results (refer 5.1.2). The PPO model was thus selected as the best performing model and the subsequent analysis of the impact of predictor variables was based on the estimates and class predictions from the PPO model (Table 5 & 6 and Figs. 2-5).

4.2. Comparison of the model performance of the PPO models for the upstream and downstream datasets

Based on Pseudo-R^2, AICc, and average train set accuracy, the downstream PPO models performed better on the train data. However, the upstream models outperformed the downstream models on the test set prediction accuracy. The lowest AICc and thereby a better fit, was observed for downstream model with interaction (1670.492). In contrast, the downstream model without interaction reported the highest Pseudo-R^2 (0.054) and average train set accuracy (0.493). The upstream model with interaction had the highest accuracy on the test set (0.544) followed by the model without interaction (0.522). Models across both the datasets showed reasonable sensitivity and specificity for the injury levels. It can be concluded that the downstream models performed better than upstream models in learning the patterns in the data, whereas the upstream models showed better performance on the unseen data. The PPO models showed reasonable multiclass prediction capability.

4.3. Impact of the contributing factors on injury severity across the upstream and downstream PPO models

In the present section, statistically significant factors in the models are reported. Class prediction rates (PR) were utilized to analyze the association between contributing factors and injury levels. The impact of a particular feature in terms of its PR of the three injury levels was given by:

$$\textit{Prediction Rate of an injury level 'a' for a feature 'X'} (\%) = (\textit{Instances of 'a' in N}) * \frac{1}{N} * 100,$$

where 'N' refers to all the predicted instances in which 'X' was involved.

For example, if X has a 10 % PR for fatal crashes, it means that if X was involved in N predicted instances, 10 % of them were fatal. (Figs. 2-5) shows the PR for the contributing factors for different severity levels. For the driver and crash related factors, the PR was displayed only for the models with 'no interaction' as their corresponding PR in the respective models with interaction differed negligibly. As curve-based crashes are critical from a fatality perspective, only the impact of the factors on the PR of fatal crashes have been discussed in the following sections.

4.3.1. Impact of driver and crash causal factors.

The statistical significance and PR of driver and crash causal factors for fatal crashes are summarized below:

1. Dangerous driving, overtaking, and speeding were statistically significant across models in both the upstream and downstream datasets (Table 5 and 6), whereas 'Road infrastructure and Alignment' was found to be insignificant. The models across both the datasets revealed the two-wheeler to be significantly linked with hospitalizations. Driver age significantly affected injury severity only in the downstream dataset where driver age $\geq$ 50 years had a marginally significant association with hospitalizations. The association between driver age $\leq$ 30 years and hospitalization were marginally significant ($p = 0.059$) in models without interaction, but significant ($p = 0.046$) in models with interaction.
2. A 'Carless Turn' was the most dominant crash cause that contributed to fatal collisions across models in both datasets. Speeding (18.33 %) and dangerous driving (9.25 %) were the second highest contributors to fatal collisions respectively in the upstream and downstream models (Fig. 2: (a)-(b)).

Table 5
Estimates for the contributing factors in the Upstream (US) dataset based on the Partial Proportional Odds (PPO) Model.

	US PPO Model-No Interaction				US PPO Model-With Interaction			
Factor	**B**	**SE**	**p-value**	**OddsRatio**	**β**	**SE**	**p-value**	**Odds Ratio**
1. ***Driver and Crash Factors***								
(a) **Crash Cause**								
Careless Turn **(ref)**	–	–	–	–	–	–	–	–
Dangerous Driving	0.707	0.257	0.006	2.027	0.694	0.256	0.007	2.001
Overtaking	0.938	0.297	0.002	2.554	0.960	0.295	0.001	2.612
Road Infra	−0.001	0.322	0.998	0.999	−0.022	0.319	0.946	0.978
Speeding	0.554	0.253	0.029	1.739	0.535	0.252	0.034	1.707
(b) **Vehicle Type**								
HMV **(ref)**	–	–	–	–	–	–	–	–
LMV: Hospitalization	0.173	0.189	0.360	1.189	0.137	0.189	0.467	1.147
LMV: Grievous	0.107	0.202	0.595	1.113	0.091	0.201	0.650	1.095
Two Wheeler: Hospitalization	−0.650	0.254	0.011	0.522	−0.657	0.254	0.010	0.519
Two Wheeler: Grievous	0.200	0.266	0.452	1.221	0.188	0.265	0.480	1.206
2. ***Subject Curve Attributes***								
(a) **Radius of Subject Curve (RAD)**								
SHS (40–60 m) **(ref)**	–	–	–	–	–	–	–	–
MKS (60–80 m)	−0.147	0.199	0.460	0.863	–	–	–	–
VSS (0–40 m)	−0.127	0.182	0.485	0.881	–	–	–	–
SFS (>80 m)	−0.485	0.249	0.051	0.615	–	–	–	–
(b) **Length of Subject Curve (LEN)**								
MLC (60–90 m) **(ref)**	–	–	–	–				
LLC (>90 m)	−0.226	0.214	0.292	0.798	–	–	–	–
SLC (30–60 m)	−0.479	0.158	0.002	0.620	–	–	–	–
3. ***Indicators of Spatial Relationship***								
(a) **Curve Direction**								
Opposite **(ref)**	–	–	–	–	–	–	–	–
Same	−0.259	0.144	0.073	0.772	−0.301	0.143	0.036	0.740
(b) **Radius of Upstream Curve (USRAD)**								
MKS (60–80 m) **(ref)**	=	=	=	=	=	=	=	=
SFS (>80 m): Hospitalization	0.003	0.240	0.989	1.003	–	–	–	–
SFS (>80 m): Grievous	−0.346	0.266	0.193	0.708	–	–	–	–
VSS (0–40 m): Hospitalization	0.010	0.223	0.965	1.010	–	–	–	–
VSS (0–40 m): Grievous	−0.226	0.243	0.354	0.798	–	–	–	–
SHS (40–60 m): Hospitalization	−0.027	0.194	0.891	0.974	–	–	–	–
SHS (40–60): Grievous	−0.407	0.216	0.059	0.666	–	–	–	–
(c) **Length of Upstream Curve (USLEN)**								
LLC (>90 m) **(ref)**	–	–	–	–	–	–	–	–
MLC (60–90 m): Hospitalization	−0.098	0.266	0.713	0.907	–	–	–	–
MLC (60–90 m): Grievous	−0.889	0.354	0.012	0.411	–	–	–	–
SLC (30–60 m): Hospitalization	−0.260	0.285	0.361	0.771	–	–	–	–
SLC (30–60 m): Grievous	−0.991	0.367	0.00	0.371	–	–	–	–
4. ***Intensity Of Sharpness***								
(a) **Approach Curve (USRAD_USLEN)**								
LLC-MKS **(ref)**	–	–	–	–	–	–	–	–
MLC-SFS	–	–	–	–	−0.028	0.252	0.910	0.972
SLC-VSS	–	–	–	–	−0.286	0.228	0.210	0.751
MLC-SHS	–	–	–	–	−0.100	0.230	0.664	0.905
SLC-MKS	–	–	–	–	0.015	0.243	0.951	1.015
(b) **Subject Curve (RAD_LEN)**								
MLC-SHS **(ref)**	–	–	–	–	–	–	–	–
LLC-MKS	–	–	–	–	0.013	0.202	0.948	1.013
SLC-VSS	–	–	–	–	−0.465	0.174	0.008	0.628
SLC-MKS	–	–	–	–	−0.239	0.201	0.235	0.788
Intercepts								
(Intercept): Hospitalization	−0.333	0.389	0.392	–	−0.472	0.328	0.151	–
(Intercept): Grievous	1.805	0.458	0.000	–	0.699	0.332	0.035	–

Note:
Ref: Reference Subcategory.
Hospitalization: Injury needing hospitalization.
Road Infra: Road Infrastructure and Alignment.
Overtaking: Overtaking or Lane change.
β: Estimate.
SE: Standard Error.

Table 6
Estimates for the contributing factors in the Downstream (DS) dataset based on the Partial Proportional Odds (PPO) Model.

	DS PPO Model-No Interaction				DS PPO Model-Interaction Model			
Factor	B	SE	p-value	Odds ratio	β	SE	p-value	Odds ratio
1. ***Driver and Crash Factors***								
(a) **Crash Cause**								
Careless Turn **(ref)**	–	–	–	–	–	–	–	–
Dangerous Driving	1.046	0.259	0.000	2.846	1.044	0.257	0.000	2.841
Overtaking	1.058	0.290	0.000	2.880	1.062	0.288	0.000	2.892
Road Infra	0.456	0.319	0.154	1.577	0.476	0.318	0.134	1.610
Speeding	1.085	0.255	0.000	2.959	1.099	0.254	0.000	3.001
(b) **Vehicle Type**								
HMV **(ref)**	–	–	–	–	–	–	–	–
LMV: Hospitalization	−0.051	0.189	0.787	0.950	−0.023	0.189	0.902	0.977
LMV: Grievous	0.036	0.202	0.857	1.037	0.072	0.202	0.723	1.074
Two Wheeler: Hospitalization	−0.857	0.259	0.001	0.425	−0.793	0.258	0.002	0.453
Two Wheeler: Grievous	0.261	0.283	0.356	1.299	0.354	0.281	0.208	1.425
(c) **Driver Age**								
≤30 years: Hospitalization	0.316	0.158	0.046	1.372	0.299	0.158	0.059	1.348
≤30 years: Grievous	0.143	0.174	0.413	1.153	0.109	0.174	0.531	1.115
30_to_50 **(ref)**	–	–	–	–	–	–	–	–
≥50 years: Hospitalization	0.490	0.253	0.053	1.632	0.478	0.253	0.058	1.614
≥50 years: Grievous	−0.267	0.265	0.315	0.766	−0.270	0.265	0.308	0.763
2. ***Subject Curve Attributes***								
(a) **Radius Of Subject Curve (RAD)**								
SHS (40–60 m) **(ref)**	–	–	–	–	–	–	–	–
VSS (0–40 m): Hospitalization	−0.158	0.192	0.411	0.854	–	–	–	–
VSS (0–40 m): Grievous	−0.402	0.201	0.045	0.669	–	–	–	–
MKS (60–80 m): Hospitalization	−0.096	0.213	0.650	0.908	–	–	–	–
MKS (60–80 m): Grievous	−0.498	0.232	0.032	0.608	–	–	–	–
SFS (>80 m): Hospitalization	−0.426	0.291	0.144	0.653	–	–	–	–
SFS (>80 m): Grievous	−0.411	0.315	0.192	0.663	–	–	–	–
(b) **Length Of Subject Curve (LEN)**								
MLC (60–90 m) **(ref)**	–	–	–	–	–	–	–	–
LLC (>90 m)	−0.070	0.208	0.736	0.932	–	–	–	–
SLC (30–60 m)	−0.538	0.162	0.001	0.584	–	–	–	–
3. I***ndicators Of Spatial Relationship***								
(a) **Curve Direction**								
Opposite **(ref)**	–	–	–	–	–	–	–	–
Same	−0.144	0.147	0.327	0.866	−0.208	0.144	0.149	0.812
(b) **Radius of Downstream Curve (DSRAD)**								
VSS (0–40 m) **(ref)**	–	–	–	–	–	–	–	–
MKS (60–80 m)	0.071	0.188	0.704	1.074	–	–	–	–
SHS (40–60 m)	0.070	0.174	0.687	1.073	–	–	–	–
SFS (>80 m)	0.046	0.246	0.851	1.047	–	–	–	–
(c) **Length of Downstream Curve (DS LEN)**								
MLC (60–90 m) **(ref)**	–	–	–	–	–	–	–	–
SLC (30–60 m)	−0.163	0.155	0.292	0.849	–	–	–	–
LLC (>90 m)	0.067	0.223	0.765	1.069	–	–	–	–
4. ***Intensity of Sharpness***								
(a) **Approach Curve (DSRAD_DSLEN)**								
MLC-SHS **(ref)**	–	–	–	–	–	–	–	–
SLC-MKS	–	–	–	–	−0.238	0.191	0.213	0.788
SLC-VSS	–	–	–	–	−0.139	0.180	0.440	0.870
LLC-SFS	–	–	–	–	0.219	0.238	0.357	1.245
(b) **Subject Curve (RAD_LEN)**								
MLC-SHS **(ref)**	–	–	–	–	–	–	–	–
LLC-MKS	–	–	–	–	−0.140	0.199	0.484	0.870
SLC-VSS	–	–	–	–	−0.636	0.181	0.000	0.529
SLC-MKS	–	–	–	–	−0.338	0.189	0.074	0.713
Intercepts								
(Intercept): Hospitalization	−0.899	0.339	0.008	–	−0.844	0.317	0.008	–
(Intercept): Grievous	0.461	0.343	0.179	–	0.381	0.319	0.233	–

Note:
ref: Reference Subcategory.
Hospitalization: Injury needing hospitalization.

Road Infra: Road Infrastructure and Alignment.
Overtaking: Overtaking or Lane change.
β: Estimate.
SE: Standard Error.
Less_eq_30 (≤30 years): ≤30 years.
30_to_50: 30–50 years.
Greater_eq_50 (≥50 years): ≥50 years.

Table 7
Predictive Performance of the Developed Models.

–	Upstream Model-No Interaction			Upstream Model-With Interaction			Downstream Model-No Interaction			Downstream Model-With Interaction		
	MULTI	ORDI	PPO	MULTI	ORDI	PPO	MULTI	ORDI	PPO	MULTI	ORDI	PPO
1. Likelihood Ratio Test (LRT)-(p-value)												
	0.000	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
2. Pseudo-R^2												
	0.059	0.0307	0.046	**0.049**	0.0265	0.035	**0.063**	0.035	0.054	**0.062**	0.032	0.049
3. Adjusted Akaike Information Criteria AICc												
	1682.769	1700.174	**1682.227**	1686.123	1701.042	**1683.894**	1701.229	1701.902	**1676.840**	1685.312	1697.200	**1670.492**
4. Omnibus Brant Test (OBT) –(p-value)												
	–	**0.000**	–	–	**0.000**	–	–	**0.000**	–	–	**0.000**	–
5. Class-wise Prediction Metrics for Training Set												
(a). Accuracy (Overall)												
	0.485	0.461	0.468	**0.483**	0.465	0.474	0.486	0.469	**0.493**	**0.496**	0.469	0.485
(b). Sensitivity (Class-wise)												
I	0.729	0.828	0.717	0.744	0.837	0.762	0.705	0.853	0.758	0.749	0.861	0.776
G	0.374	**0.000**	0.265	0.344	**0.000**	0.248	0.315	**0.000**	0.263	0.302	**0.000**	0.246
F	0.243	0.392	0.308	0.243	0.392	0.282	0.329	0.374	0.329	0.311	0.360	0.293
(c). Specificity (Class-wise)												
I	0.477	0.337	0.440	0.455	0.322	0.418	0.467	0.311	0.443	0.447	0.286	0.403
G	0.826	**1.000**	0.869	0.842	**1.000**	0.880	0.847	**1.000**	0.870	0.873	**1.000**	0.889
F	0.877	0.783	0.838	0.875	0.801	0.854	0.860	0.811	0.867	0.863	0.830	0.869
6. Class-wise Prediction Metrics for Test Set												
(a). Accuracy (Overall)												
	0.507	0.478	**0.522**	0.507	0.486	**0.544**	**0.457**	0.464	0.449	0.457	0.493	**0.493**
(b). Sensitivity (Class-wise)												
I	0.793	0.793	0.532	0.810	0.828	0.553	0.678	0.814	0.627	0.695	0.864	0.763
G	0.325	**0.000**	0.433	0.300	**0.000**	0.480	0.225	**0.000**	0.300	0.225	**0.000**	0.250
F	0.275	0.500	0.586	0.275	0.475	0.571	0.359	0.410	0.333	0.333	0.436	0.333
(c). Specificity (Class-wise)												
I	0.500	0.400	0.729	0.400	0.413	0.793	0.418	0.291	0.430	0.456	0.279	0.405
G	0.847	**1.000**	0.750	0.888	**1.000**	0.752	0.857	**1.000**	0.827	0.847	**1.000**	0.867
F	0.867	0.755	0.789	0.908	0.755	0.782	0.849	0.818	0.859	0.828	0.869	0.899

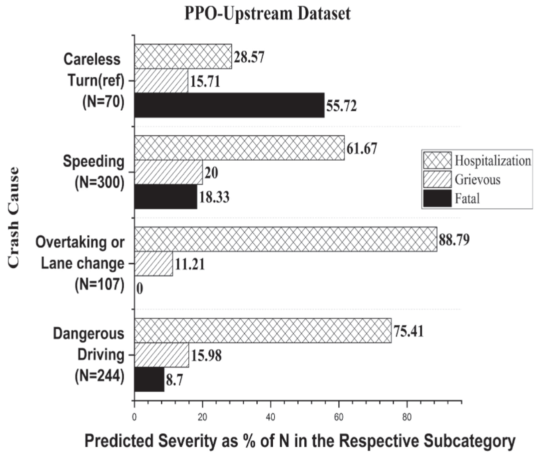
Note:
F = Fatal.
G = Grievous.
I = Injury needing hospitalization.
MULTI = Multinomial Logit Model.
ORDI = Ordinal Logit Model.
PPO = Partial Proportional Odds Model.

3. The models in both datasets predicted that HMVs had the highest fatality and hospitalization rates, whereas Two-Wheelers had the highest rate of grievous crashes (Fig. 2: (c)-(d)).
4. Additionally, the models predicted middle aged drivers to be associated with the highest rate (24.44 %) of fatal crashes followed by old (20.48 %) and young (20.29 %) drivers (Fig. 2: (e)).

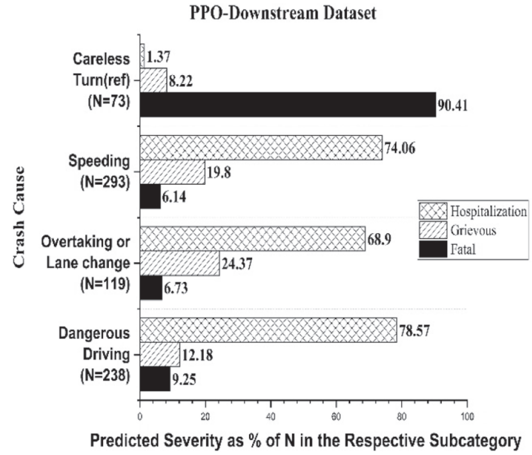
4.3.2. *Impact of curve attributes (radius and length) of the subject curve.*

The statistical significance and PR of the impact of curve attributes (radius and length) of the subject curve for fatal crashes are summarized below:

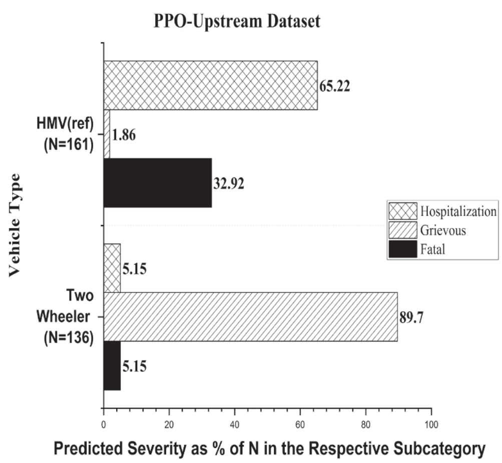
1. A subject curve with Radius > 80 m was found to be significant in the upstream dataset. Further, the PPO model in the downstream predicted a subject curve with a very sharp (Radius ≤ 40 m) or marked curvature (Radius:60–80 m) to be significantly associated with grievous crashes. Both datasets identified a short subject curve (30–60 m) as significant.
2. Fatal crash prediction rates for the subject curve's radius differed between models in the two datasets. The upstream dataset (no-interaction) model predicted curves with (Radius > 80 m) to have the highest fatality rate (36.23 %), while the downstream dataset model predicted it for curves with a very sharp curvature (38.97 %) and marked curvature (15.07 %) (Fig. 3: (a)-(b)).



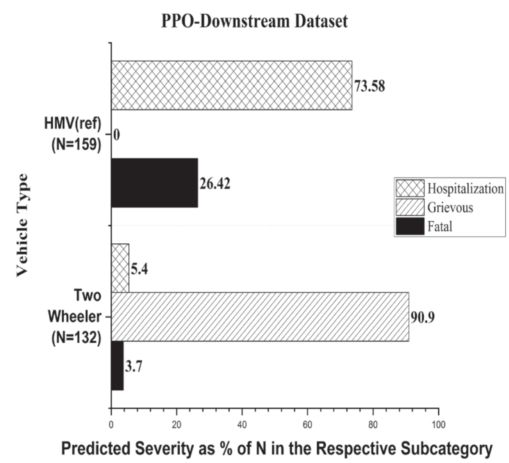
(a) Prediction Rate-Crash Cause (US)



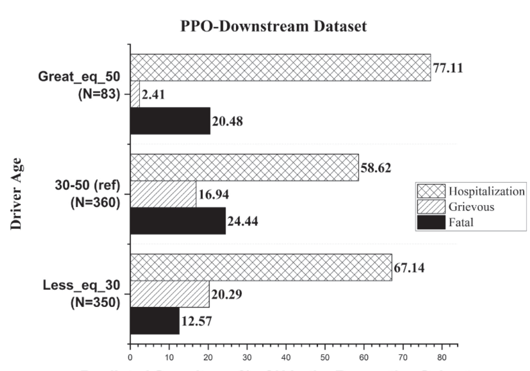
(b) Prediction Rate-Crash Cause (DS)



(c) Prediction Rate- Vehicle Type (US)

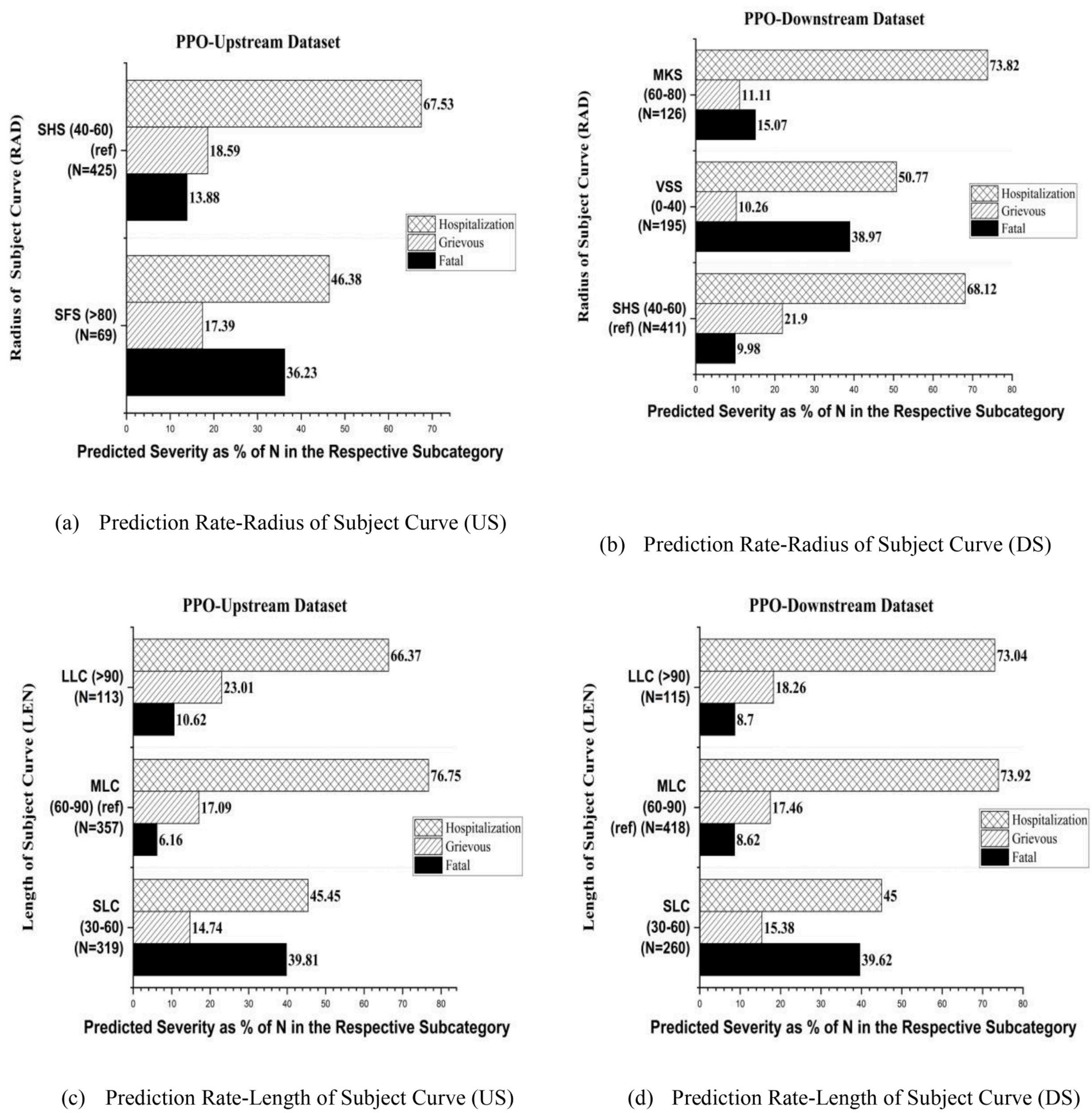


(d) Prediction Rate- Vehicle Type (DS)



(e) Prediction Rate- Driver Age (DS)

Fig. 2. Prediction Rate for Crash Cause, Vehicle Type and Driver Age for the Upstream (US) and Downstream (DS) Models.



(a) Prediction Rate-Radius of Subject Curve (US)

(b) Prediction Rate-Radius of Subject Curve (DS)

(c) Prediction Rate-Length of Subject Curve (US)

(d) Prediction Rate-Length of Subject Curve (DS)

Fig. 3. Prediction Rate for Curve Attributes (radius and length) of the Subject Curve for the Upstream (US) and Downstream (DS) Models.

3. Given a crash, a short curve resulted in the highest rate of fatalities followed by a long (length > 90 m) and medium length (length: 60–90 m) curve across models in both the datasets (Fig. 3: (c)-(d)).

4.3.3. Impact of the indicators of spatial relationship.

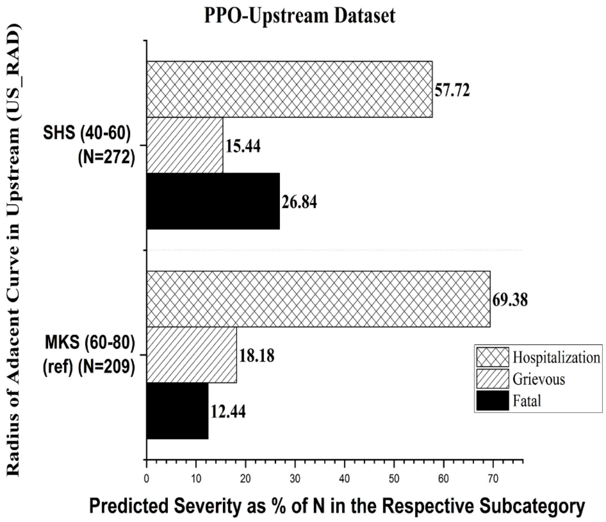
The statistical significance and PR of the impact of the indicators of spatial relationship for fatal crashes are summarized below:

1. All the indicators of spatial relationship were found to be insignificant for the models in the downstream dataset (Table 6). A sharp approach curve was linked with grievous injuries in the upstream dataset (Table 5), although only marginally ($p = 0.059$). Additionally, a short length of the approach curve in the upstream dataset contributed significantly ($p = 0.007$) to grievous injuries. For the factor 'curve direction', the relative direction of turn for adjacent curves in the 'same' direction was significant in the models across the upstream interaction and no-interaction models ($p = 0.073$ and $p = 0.036$ respectively).
2. Given a crash on the subject curve, a sharp curve upstream was most likely to result in a fatal crash (26.84 %) followed by a curve with a marked curvature (12.44 %) (Fig. 4: (a)).
3. A short upstream curve was the most significant contributor to the rate of fatalities (33.14 %), given a crash on the subject curve (Fig. 4: (b)).
4. The models indicated successive curves turning in the same direction to result in the highest fatality rate (32.84 %) (Fig. 4: (c)).

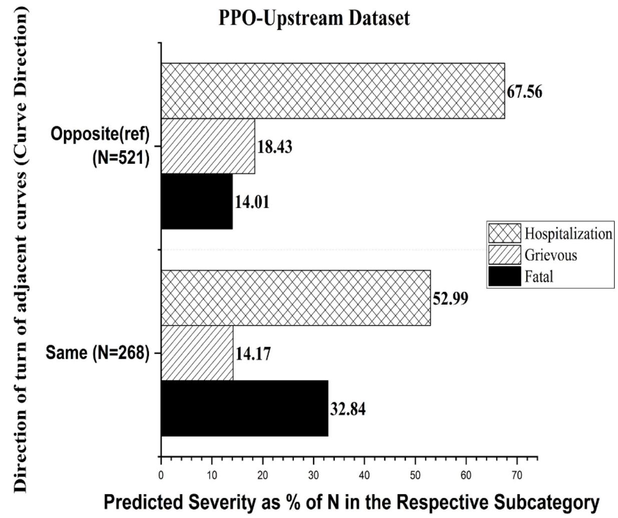
4.3.4. Impact of the intensity of sharpness for the subject and approach curve.

The statistical significance and PR of the impact of the intensity of sharpness for the subject and approach curve for fatal crashes are summarized below:

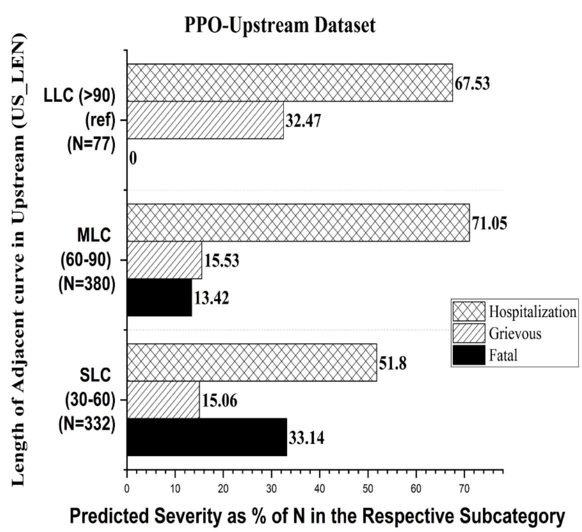
1. The intensity of sharpness of the approach curve was predicted to be insignificant across the datasets in both directions (Tables 5 and 6). A short subject curve with very sharp curvature (SLC-VSS) was



(a) Prediction Rate-Radius of Adjacent Curve (US)



(b) Prediction Rate-Curve Direction (US)



(c) Prediction Rate-Length of Adjacent Curve (US)

Fig. 4. Prediction Rate for the Curve Attributes of the Adjacent Approach curve (radius and length) and its Relative Direction of Turn for the Upstream (US) Models.

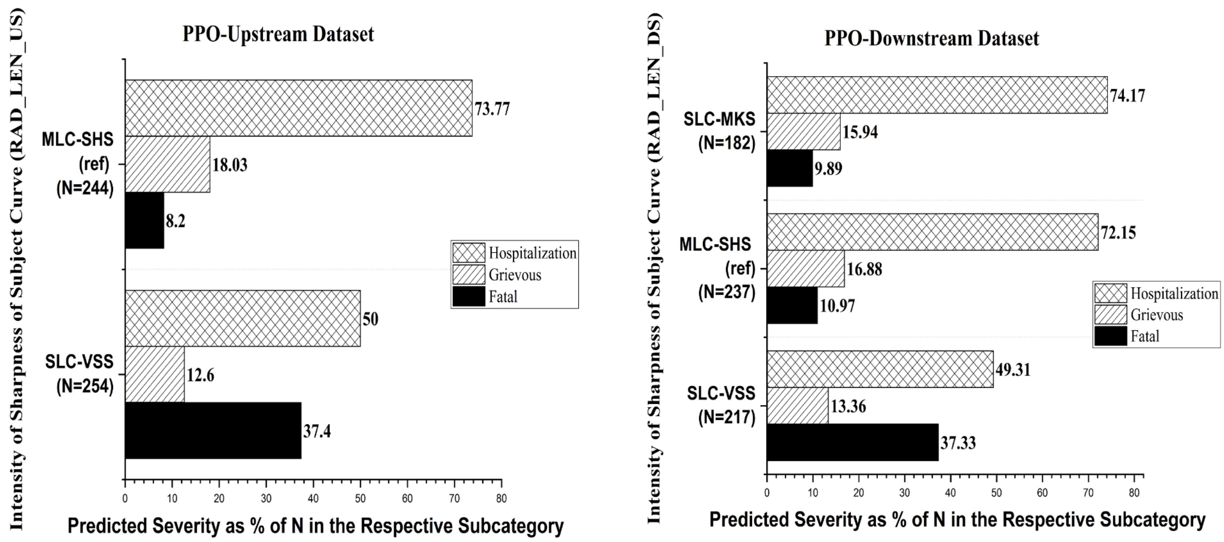
invariably found to be significant. Furthermore, the downstream PPO model predicted a short length curve with marked sharpness (SLC-MKS) to be marginally significant ($p = 0.074$) in its association with severity.

2. The highest fatality rate was predicted by models across both datasets for a crash on a short length subject curve with a very sharp curvature (SLC-VSS), followed by a medium length curve with a sharp curvature (MLC-SHS). (Fig. 5: (a)-(b)).

5. Discussion

5.1. Effect of crash and driver related factors on injury severity.

Careless turn resulted in at least 3 times more fatalities compared to other major crash causes. A careless turn on mountainous curves can most likely result in a ROR crash if the driver fails to anticipate or assess the curve's sharpness, thereby losing control. Running off the road in absence of safety measures such as guardrails or parapets on the valley side can cause the vehicle to fall from substantial heights increasing the risk of fatality especially for HMVs owed to their momentum. The lack of protection in two-wheelers increases the risk of severe crashes. Speeding leads to loss of control on curves and a higher impact in the event of a collision increasing the risk of serious injuries by 1.2–1.7 times. Due to limited visibility at curves, drivers attempting to overtake slower heavy vehicles can meet severe head-on collisions, explaining the significant involvement of overtaking in grievous crashes. The significant involvement of middle (30–50 years) and old (≥50 years) aged drivers in fatal and grievous crashes can be attributed to the disadvantage in terms of rection time and visibility for older drivers whereas for middle aged drivers it would be overconfidence resulting in reduced vigilance



(a) Prediction Rate-Intensity of Sharpness of Subject Curve (US)

(b) Prediction Rate-Intensity of Sharpness of Subject Curve (DS)

Fig. 5. Prediction Rate for the Intensity of Sharpness of The Subject Curve for the Upstream (US) and Downstream (DS) Models.

(Ahmadi et al., 2020; Azmeri Khan et al. (2023); Chen et al., 2022; Liu et al., 2021; Montella et al., 2021; Zhou and Chin, 2019).

5.2. Effect of the curve attributes of the subject curve on injury severity

The upstream and downstream PPO models reported contradictory results for fatal crashes with the upstream models revealing relatively flatter subject curves (radius $>$ 80 m) to most likely result in fatality whereas the downstream models identified relatively sharper curves (radius $\leq$ 40 m) as riskier. The curvature characteristics of the leading subject curves are likely to differ across the upstream and downstream datasets. Furthermore, despite serial subject curves being jointly shared by the two datasets, the difference in the curvature characteristics of their corresponding approach curves in a given direction combined with effect modification due to simultaneous presence of other factors could have resulted in contradictory inferences from the corresponding models. To resolve the issue three options were considered: (1) percentage of fatal crashes for the two contradicting subcategories in the training dataset, (2) findings of past research on the impact of curve radius and (3) comparison of the PPO model fit for the train data. These three would collectively help draw conclusions for the observed trend in the models based on descriptive statistics, agreement in past studies on the impact of curve radius, and model fit as an indicator of their ability to capture data patterns and associate them with injury severity. Very sharp (radius $\leq$ 40 m)) subject curves consistently outnumbered those with significant curvature (radius $>$ 80 m) across the upstream and downstream datasets in terms of fatal crash percentage respectively by 0.72 % and 7.41 %. Further, the past studies also predicted a reduction in curve sharpness to lower the crash rates (Gooch et al., 2016; Khan et al., 2013; Kronprasert et al., 2021). The downstream models not only outperformed the upstream models with a lower AICc (1676.84 vs 1682.22 respectively), but also in terms of average overall accuracy, sensitivity and specificity (for fatal crashes) in the train set predictions by 5.34 %,6.82 % and 3.46 % respectively. This conclusively provided evidence for very sharp (radius $\leq$ 40 m) subject curves to increase the risk of fatality. Both short and long subject curves significantly increased the relative risk of fatal and grievous crashes, respectively by 3(or more) and (1.1–1.5) times. Short curves, especially sharp ones, introduce the curvature abruptly, significantly reducing the effective reaction time to exert the necessary control. Longer curves, on the other hand, may encourage speeding and other risky maneuvers despite introducing the curvature gradually as reported in the past (Khan et al., 2013; Kronprasert et al., 2021).

5.3. Effect of the indicators of spatial relationship on injury severity

A sharp curve (SHS) upstream of the subject curve increased the fatality risk by 2 times, contrary to past studies, that found a sharp nearby curve to lower the crash rate (Elvik, 2019; Gooch et al., 2016; Hamilton et al., 2019). Most of the studies on spatial relationships have been limited to crash rate estimation. However, Hosseinpour et al., (2014) found that explanatory factors can have different or even opposite effects on crash rate and severity. Sharp curves, especially those separated by a small tangent, induce instability due to curve forces, are visually demanding (Easa and Ganguly, 2005), making it hard to anticipate the subsequent curve or spot an oncoming vehicle and thereby make it difficult for the driver to perform appropriate maneuvers.

A short approach curve increasing risk of fatality by 2.5 times, can be attributed to the increased workload and a reduced opportunity to anticipate the curvature or presence of the upcoming curve. Longer curves resulted in twice as many serious crashes as they can encourage speeding and can cause drivers to lose control as they approach a sharp turn, especially if its sharpness was underestimated.

Successive curves turning opposite reduced the fatality risk. The findings are consistent with those for reverse curves as drivers are expected to be more cautious and avoid speeding (Kronprasert et al., 2021). IRC 52:2019 recommends avoiding consecutive curves turning in the same direction separated by small tangents (referred to as broken-back curves) as they can instill a false sense of consistency, encouraging higher speeds that may have resulted in relatively 2.3 times higher fatalities compared to the curves turning relatively opposite(Tiwari and Mohan, 2018).

5.4. Effect of the intensity of sharpness for the approach and subject curve on injury severity

The overall sharpness of the curve experienced by the driver is likely to increase for a subject curve characterized by a short length and a very sharp curvature (SLC-VSS) thereby potentially increasing the workload to retain vehicular control that resulted in 3 times fatalities in such a situation. The significant association of the subject curve's intensity of sharpness with severity indicated that the radius and length of the curve together shape its overall perceived sharpness. The inference from the intensity of sharpness contradicts the upstream PPO model, which revealed that flatter subject curves (Rad > 80 m) were riskier instead of very sharp curve (Rad ≤ 40 m). The intensity of sharpness highlights that assessing a curve's sharpness as combination of its length and radius increases the reliability of the results. However, the insignificant estimates for the intensity of sharpness for the approach curve were unexpected and a similar analysis with a larger sample size would provide a more in-depth picture.

5.5. Discussion on the overall performance of the crash severity models

In terms of the goodness of fit (GOF) measures, the PPO models in the downstream dataset outperformed their upstream counterparts, with lower AIC, higher pseudo-R2 and more accurate train set predictions. However, the downstream models produced insignificant estimates for the indicators of spatial relationship between the subject and approach curve. In contrast, upstream models yielded significant estimates for spatial relationship indicators and higher test-set accuracy. This difference may indicate that the driver at a given point of time could have been influenced by the approach curve in only one direction. The PPO models for both the datasets showed a reasonable multiclass prediction capability unlike ORDI models. This shows that a severity prediction model evaluated solely on the GOF and overall accuracy, without regard for its class-specific performance, may appear to perform well while in reality it could have omitted a significant amount of relevant information.

6. Conclusion

The study contributes additional evidence in support of spatial relationship between adjacent curves and their interactive relationship in influencing the overall severity, reflecting the short-term driver expectancy. Further, intensity of sharpness for the subject curve was found significant highlighting the fact that the overall perception of a curve's sharpness with a given radius can be simultaneously influenced by its length. Additionally, the spatial relationship was also modeled using adjacent curve's features and their relative turn direction, which were found to be significant. The human element and the geometric design characteristics were simultaneously modelled providing a more in-depth analysis. Model validation based on a train-test split employing stratified sampling ensured proportionate representation of the severity levels across the two datasets. The results can help safety professionals in identifying high risk scenarios and planning appropriate safety enhancement strategies.

A significant spatial relationship between adjacent curves was found to exist supporting their interactive relationship in influencing the severity outcome. A sharp curve (radius:40–60 m) adjacent to a given curve increased the risk of fatality by 2.16 times with a similar increase (by 2.5 times or more) observed for a short (length:30–60 m) adjacent curve. Adjacent curves turning in the same direction increased the risk of fatality by 2.34 times. For the subject curve an increase in curvature specifically with radius ≤ 40 m and a short length (30–60 m) increased the risk of fatality for both by 3 times in general (individually). Additionally subject curves characterized by a short length with very sharp curvature resulted in 3–4 times higher fatalities. The identified high-risk scenarios can help to design efficient pre-warning systems along with more informative navigation systems especially from a safety viewpoint. Under such risky circumstances, making erroneous maneuvers like careless turns, over-speeding along with deficient protective infrastructure measures were found to further aggravate the risk. Therefore, appropriate protective measures like guardrails etc. and no overtaking signs on curves can be installed along with ensuring more consistent highway designs. The higher risk associated with HMVs demands more attention for routes dominated by such traffic. Specifically, age-based education programs and interactive driving assistant systems can further reduce the risk. More detailed crash reporting and efficient crash data management systems are crucial for continuous safety assessment, especially in developing countries.

6.1. Limitations and future scope

Unavailability of design drawings and road inventory data limited the number of geometric factors that could be used. The results of clustering are essentially data driven, which are likely to change as per the given dataset. Further, a crash record for only 6 years was available. The crash description in terms of the approach direction for the curve was not available which would have further enhanced the reliability of the results. The statistical models employed in the study were 'fixed models' and generally have their own limitations**.** In future, machine learning classification models can be employed which are not bound by the distribution of the underlying data and assumptions of the statistical models. However, their interpretability remains a major concern. Additional data mining techniques that can extract interactive relationships between the contributing factors can also be used in future. Even though this study did not account for temporal inconsistencies in developing the severity models, however due to negligible changes in safety infrastructure during the given period to affect crash patterns, insignificant change in travel patterns, consistent increase in overall crash risk and personal vehicle ownership, there might be less likelihood of such temporal inconsistencies to occur. A larger and more comprehensive data set would enhance the reliability of the results. Driver simulator-based experiments can be designed in the future based on the identified high-risk scenarios in the current study, with more driver characteristics ensuring a more in-depth safety analysis of curves in sequence.

Funding source

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

CRediT authorship contribution statement

Deepak Awasthi: Writing – original draft, Methodology, Formal analysis, Conceptualization. **Raman Parti:** Writing – original draft. **Kirti Mahajan:** Writing – original draft, Methodology, Formal analysis, Conceptualization.

Declaration of competing interest

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

Data availability

Data will be made available on request.

References

Ahmadi, A., Jahangiri, A., Berardi, V., Machiani, S.G., 2020. Crash severity analysis of rear-end crashes in California using statistical and machine learning classification

methods. J. Transp. Saf. Secur. 12, 522–546. https://doi.org/10.1080/19439962.2018.1505793.
Alrejjal, A., Farid, A., Ksaibati, K., 2022. Investigating factors influencing rollover crash risk on mountainous interstates. J. Safety Res. 80, 391–398. https://doi.org/10.1016/j.jsr.2021.12.020.
Al-Sahili, K., Dwaikat, M., 2019. Modeling geometric design consistency and road safety for two-lane rural highways in the west bank. Palestine. Arab. J. Sci. Eng. 44, 4895–4909. https://doi.org/10.1007/s13369-018-3610-7.
Al-Sahili, K., Dwaikat, M., Abu-Eisheh, S., Alhajyaseen, W., 2018. Effectiveness of consistency measures in crash prediction models for two-lane highways in Palestine. Arab. J. Sci. Eng. 43, 5645–5656. https://doi.org/10.1007/s13369-018-3285-0.
Azmeri Khan, S., Yasmin, S., Mazharul Haque, M., 2023. Effects of design consistency measures and roadside hazard types on run-off-road crash severity: application of random parameters hierarchical ordered probit model. Anal. Methods Accid. Res. 40, 100300 https://doi.org/10.1016/j.amar.2023.100300.
Bíl, M., Andrášik, R., Sedoník, J., Cícha, V., 2018. ROCA – an ArcGIS toolbox for road alignment identification and horizontal curve radii computation. PLoS One 13, e0208407.
Chakraborty, M., Gates, T.J., 2023. Relationship between horizontal curve geometry and single-vehicle crash occurrence on rural secondary highways. Transp. Res. Rec. J. Transp. Res. Board. https://doi.org/10.1177/03611981231208901.
Chaudhari, P.M., Goyani, J., Arkatkar, S., Joshi, G., Easa, S.M., 2022. Design consistency evaluation of two-lane rural highways in hilly terrains. Transp. Res. Procedia 62, 75–82. https://doi.org/10.1016/j.trpro.2022.02.010.
Chen, Y., Luo, R., Yang, H., King, M., Shi, Q., 2020. Applying latent class analysis to investigate rural highway single-vehicle fatal crashes in China. Accid. Anal. Prev. 148, 105840 https://doi.org/10.1016/j.aap.2020.105840.
Chen, Y., Luo, R., King, M., Shi, Q., He, J., Hu, Z., 2022. Spatiotemporal analysis of crash severity on rural highway: A case study in Anhui China. Accid. Anal. Prev. 165, 106538 https://doi.org/10.1016/j.aap.2021.106538.
Chen, Z., Wen, H., Zhu, Q., Zhao, S., 2023. Severity analysis of multi-truck crashes on mountain freeways using a mixed logit model. Sustainability 15, 6499. https://doi.org/10.3390/su15086499.
Deng, Z., Ivan, J., Gårder, P., 2006. Analysis of factors affecting the severity of head-on crashes: two-lane rural highways in connecticut. Transp. Res. Rec. J. Transp. Res. Board 1953, 137–146. https://doi.org/10.3141/1953-16.
Easa, S., Ganguly, C., 2005. Modeling driver visual demand on complex horizontal alignments. J. Transp. Eng. 131, 583–590. https://doi.org/10.1061/(ASCE)0733-947X(2005)131:8(583).
Elvik, R., 2019. The more (sharp) curves, the lower the risk. Accid. Anal. Prev. 133, 105322 https://doi.org/10.1016/j.aap.2019.105322.
Ferreira-Vanegas, C.M., Vélez, J.I., García-Llinás, G.A., 2022. Analytical methods and determinants of frequency and severity of road accidents: a 20-year systematic literature review. J. Adv. Transp. 2022, 1–17. https://doi.org/10.1155/2022/7239464.
Findley, D.J., Hummer, J.E., Rasdorf, W., Zegeer, C.V., Fowler, T.J., 2012. Modeling the impact of spatial relationships on horizontal curve safety. Accid. Anal. Prev. 45, 296–304. https://doi.org/10.1016/j.aap.2011.07.018.
García, A., Llopis-Castelló, D., Camacho-Torregrosa, F.J., Pérez-Zuriaga, A.M., 2013. New consistency index based on inertial operating speed. Transp. Res. Rec. J. Transp. Res. Board 2391, 105–112. https://doi.org/10.3141/2391-10.
Gooch, J.P., Gayah, V.V., Donnell, E.T., 2016. Quantifying the safety effects of horizontal curves on two-way, two-lane rural roads. Accid. Anal. Prev. 92, 71–81. https://doi.org/10.1016/j.aap.2016.03.024.
Gooch, J.P., Gayah, V.V., Donnell, E.T., 2018. Safety performance functions for horizontal curves and tangents on two lane, two way rural roads. Accid. Anal. Prev. 120, 28–37. https://doi.org/10.1016/j.aap.2018.07.030.
Hamilton, I., Himes, S., Porter, R.J., Donnell, E., 2019. Safety evaluation of horizontal alignment design consistency on rural two-lane highways. Transp. Res. Rec. J. Transp. Res. Board 2673, 628–636. https://doi.org/10.1177/0361198119829414.
Hosseinpour, M., Yahaya, A.S., Sadullah, A.F., 2014. Exploring the effects of roadway characteristics on the frequency and severity of head-on crashes: case studies from Malaysian Federal Roads. Accid. Anal. Prev. 62, 209–222. https://doi.org/10.1016/j.aap.2013.10.001.
Ikotun, A.M., Ezugwu, A.E., Abualigah, L., Abuhaija, B., Heming, J., 2023. K-means clustering algorithms: a comprehensive review, variants analysis, and advances in the era of big data. Inf. Sci. (ny) 622, 178–210. https://doi.org/10.1016/j.ins.2022.11.139.
Indian Roads Congress, 2019. IRC: 52-2019 Guidelines for the Alignment Survey and Geometric Design of Hill Roads (Third Revision).
Iranitalab, A., Khattak, A., 2017. Comparison of four statistical and machine learning methods for crash severity prediction. Accid. Anal. Prev. 108, 27–36. https://doi.org/10.1016/j.aap.2017.08.008.
Jacob, A., Anjaneyulu, M.V.L.R., 2013. Operating speed of different classes of vehicles at horizontal curves on two-lane rural highways. J. Transp. Eng. 139, 287–294. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000503.
Jeong, H., Jang, Y., Bowman, P.J., Masoud, N., 2018. Classification of motor vehicle crash injury severity: A hybrid approach for imbalanced data. Accid. Anal. Prev. 120, 250–261. https://doi.org/10.1016/j.aap.2018.08.025.
Jou, R.-C., Chao, M.-C., 2022. An analysis of the novice motorcyclist crashes in Taiwan. Traffic Inj. Prev. 23, 140–145. https://doi.org/10.1080/15389588.2022.2026937.
Khan, S.A., Afghari, A.P., Yasmin, S., Haque, M.M., 2023. Effects of design consistency on run-off-road crashes: an application of a Random Parameters Negative Binomial Lindley model. Accid. Anal. Prev. 186, 107042 https://doi.org/10.1016/j.aap.2023.107042.
Khan, G., Bill, A.R., Chitturi, M.V., Noyce, D.A., 2013. Safety evaluation of horizontal curves on rural undivided roads. Transp. Res. Rec. J. Transp. Res. Board 2386, 147–157. https://doi.org/10.3141/2386-17.
Kim, H., Kim, J.-T., Shin, S., Lee, H., Lim, J., 2022. Prediction of run-off road crash severity in South Korea's highway through tree augmented naïve bayes learning. Appl. Sci. 12, 1120. https://doi.org/10.3390/app12031120.
Kronprasert, N., Boontan, K., Kanha, P., 2021. Crash prediction models for horizontal curve segments on two-lane rural roads in Thailand. Sustainability 13, 9011. https://doi.org/10.3390/su13169011.
Li, J., Guo, F., Zhou, Y., Yang, W., Ni, D., 2023. Predicting the severity of traffic accidents on mountain freeways with dynamic traffic and weather data. Transp. Saf. Environ. 5 https://doi.org/10.1093/tse/tdad001.
Li, Z., Liu, P., Wang, W., Xu, C., 2012. Using support vector machine models for crash injury severity analysis. Accid. Anal. Prev. 45, 478–486. https://doi.org/10.1016/j.aap.2011.08.016.
Liu, P., Fan (David), W., 2021. Analysis of head-on crash injury severity using a partial proportional odds model. J. Transp. Saf. Secur. 13, 714–734. https://doi.org/10.1080/19439962.2019.1667933.
Llopis-Castelló, D., Bella, F., Camacho-Torregrosa, F.J., García, A., 2018a. New consistency model based on inertial operating speed profiles for road safety evaluation. J. Transp. Eng. Part A Syst. 144 https://doi.org/10.1061/JTEPBS.0000126.
Llopis-Castelló, D., Camacho-Torregrosa, F.J., García, A., 2018b. Calibration of the inertial consistency index to assess road safety on horizontal curves of two-lane rural roads. Accid. Anal. Prev. 118, 1–10. https://doi.org/10.1016/j.aap.2018.05.014.
Llopis-Castelló, D., Camacho-Torregrosa, F.J., García, A., 2018c. Development of a global inertial consistency model to assess road safety on Spanish two-lane rural roads. Accid. Anal. Prev. 119, 138–148. https://doi.org/10.1016/j.aap.2018.07.018.
Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transp. Res. Part A Policy Pract. 44, 291–305. https://doi.org/10.1016/j.tra.2010.02.001.
Ma, J., Ren, G., Wang, S., Yu, J., Wang, L., 2022. Characterizing the effects of contributing factors on crash severity involving e-bicycles: a study based on police-reported data. Int. J. Inj. Contr. Saf. Promot. 29, 463–474. https://doi.org/10.1080/17457300.2022.2081982.
Meng, Y., 2017. Estimation of crash severity on mountainous freeways in chongqing. Math. Probl. Eng. 2017, 1–8. https://doi.org/10.1155/2017/9764309.
Ministry of Road Transport and Highways (MORT&H), 2019. Density Distribution of National Highways as on 31.03.2019. New Delhi.
Ministry of Road Transport and Highways (MoRT&H), 2023. Road accidents in India 2022.
Montella, A., Colantuoni, L., Lamberti, R., 2008. Crash prediction models for rural motorways. Transp. Res. Rec. J. Transp. Res. Board 2083, 180–189. https://doi.org/10.3141/2083-21.
Montella, A., Imbriani, L.L., 2015. Safety performance functions incorporating design consistency variables. Accid. Anal. Prev. 74, 133–144. https://doi.org/10.1016/j.aap.2014.10.019.
Montella, A., Mauriello, F., Pernetti, M., Rella Riccardi, M., 2021. Rule discovery to identify patterns contributing to overrepresentation and severity of run-off-the-road crashes. Accid. Anal. Prev. 155, 106119 https://doi.org/10.1016/j.aap.2021.106119.
Mooradian, J., Ivan, J.N., Ravishanker, N., Hu, S., 2013. Analysis of driver and passenger crash injury severity using partial proportional odds models. Accid. Anal. Prev. 58, 53–58. https://doi.org/10.1016/j.aap.2013.04.022.
Nair, S.R., Bhavathrathan, B.K., 2022. Hybrid segmentation approach to identify crash susceptible locations in large road networks. Saf. Sci. 145, 105515 https://doi.org/10.1016/j.ssci.2021.105515.
NCRB, 2023. Accidental Deaths & Suicides in India 2022, National Crime Records Bureau Ministry of Home Affairs. New Delhi.
Polus, A., Dagan, D., 1987. Model for evaluating the consistency of highway alignment. Transp. Res. Rec. 47–56.
Polus, A., Mattar-Habib, C., 2004. New consistency model for rural highways and its relationship to safety. J. Transp. Eng. 130, 286–293. https://doi.org/10.1061/(ASCE)0733-947X(2004)130:3(286).
Revlin, R., 2012. Cognition: Theory and practice. Macmillan.
Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.
Santos, K., Dias, J.P., Amado, C., 2022. A literature review of machine learning algorithms for crash injury severity prediction. J. Safety Res. 80, 254–269. https://doi.org/10.1016/j.jsr.2021.12.007.
Savolainen, P.T., Mannering, F.L., Lord, D., Quddus, M.A., 2011. The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives. Accid. Anal. Prev. 43, 1666–1676. https://doi.org/10.1016/j.aap.2011.03.025.
Song, D., Yang, X., Yang, Y., Cui, P., Zhu, G., 2023. Bivariate joint analysis of injury severity of drivers in truck-car crashes accommodating multilayer unobserved heterogeneity. Accid. Anal. Prev. 190, 107175 https://doi.org/10.1016/j.aap.2023.107175.
Sustainable Development in The Indian Himalayan Region [WWW Document], n.d. . Niti Ayog, Gov. India. URL https://www.niti.gov.in/sustainable-development-indian-himalayan-region (accessed 5.16.24).
Tiwari, G., Mohan, D., Bhalla, K., 2022. Road Safety in India : Status Report 2021, Transportation Research & Injury Prevention Programme Indian Institute of Technology Delhi.

Tiwari, G., Mohan, D., 2016. Transport Planning and Traffic Safety. CRC Press, Taylor & Francis Group, Boca Raton, London, New York, Making Cities, Roads and Vehicles Safer.

Tiwari, G., Mohan, D. (Eds.), 2018. Transport Planning and Traffic Safety. CRC Press, Indian Institute of Technology, New Delhi, India. https://doi.org/10.1201/b19737.

Torbic, D.J., Harwood, D.W., Gilmore, D.K., Pfefer, R., Neuman, T.R., Slack, K.L., Hardy, K.K., 2004. Guidance for implementation of the AASHTO strategic highway safety plan. Volume 7: A guide for reducing collisions on horizontal curves.

Vinod, H.D., 1969. Integer Programming and the Theory of Grouping Author (s): Hrishikesh D . Vinod Source: Journal of the American Statistical Association , Vol . 64 , No . 326 (Jun ., 1969), pp . Published by : Taylor & Francis , Ltd . on behalf of the American Statis. J. Am. Stat. Assoc. 64, 506–519.

Wang, Y., Xu, G., Bai, H., 2013. Operating speed based alignment consistency evaluation using driving simulator: case studies from Taigan Freeway in Jiangxi, China. Promet - Traffic - Traffico 25, 23–31. https://doi.org/10.7307/ptt.v25i1.1244.

Wen, X., Xie, Y., Jiang, L., Pu, Z., Ge, T., 2021. Applications of machine learning methods in traffic crash severity modelling: current status and future directions. Transp. Rev. 41, 855–879. https://doi.org/10.1080/01441647.2021.1954108.

Ye, F., Lord, D., 2014. Comparing three commonly used crash severity models on sample size requirements: multinomial logit, ordered probit and mixed logit models. Anal. Methods Accid. Res. 1, 72–85. https://doi.org/10.1016/j.amar.2013.03.001.

Yu, H., Li, Z., Zhang, G., Liu, P., 2019. A latent class approach for driver injury severity analysis in highway single vehicle crash considering unobserved heterogeneity and temporal influence. Anal. Methods Accid. Res. 24, 100110 https://doi.org/10.1016/j.amar.2019.100110.

Yu, M., Ma, C., Zheng, C., Chen, Z., Yang, T., 2022. Injury severity of truck-involved crashes in work zones on rural and urban highways: accounting for unobserved heterogeneity. J. Transp. Saf. Secur. 14, 83–110. https://doi.org/10.1080/19439962.2020.1726544.

Zhou, M., Chin, H.C., 2019. Factors affecting the injury severity of out-of-control single-vehicle crashes in Singapore. Accid. Anal. Prev. 124, 104–112. https://doi.org/10.1016/j.aap.2019.01.009.