Sourcing of geometric data for Safety Performance Function: A Review and future directions

Kommoju Prathyusha, Amit Agarwal, Indrajit Ghosh

Indian Institute of Technology Roorkee, India

Abstract. The paper reviews the different geometric variables needed for developing an SPF and the sources or survey methods for collecting them. In this process, the relationship between individual road geometry variables (qualitative and quantitative) and crashes is realized to exhibit mixed nature (i.e., increase or decrease). The findings of the review on road geometric variables indicated the need for further study on the interrelation between variables, understanding the influence of different variables on urban and rural roads of less importance, the inclusion of variables qualitative, which are difficult to collect, etc. From the data sources or survey methods and the type of data, we can imply that every method employed is either not able to collect all the possible geometric variables influencing crashes and their condition or expensive or time-consuming process. Many of the techniques are also difficult to extract for larger study regions. This causes restricted study sites.

keyword: Safety performance function; crash frequency analysis; crash prediction model

1 Introduction

Background: As per the WHO, 1.3 billion people die every year owing to road traffic crashes, and 20-50 million people are suffering from non-fatal injuries. It is also observed that more than 90% of the victims are from middle and low-income groups. This necessitates the development of a system that pushes towards zero crash rates.

Developing economies like India are severely affected by road crashes. These countries predominantly have heterogeneous traffic conditions, which consist of vehicles of different dynamics (like the angle of rotation of the front wheel, the height of the driver's seat, weight-to-power ratio, etc.) and driver characteristics [1, 2]. These may lead to the requirement of different geometric standards (like safe turning radius, visibility, etc.). Further, the absence of lane discipline, use of the same road space by motorized and non-motorized traffic, poor licensing, etc., are making the situation more challenging.

Typically, to analyze road crashes in different scenarios, a risk analysis can be performed using the reactive approach (post-crash analysis), and the proactive approach (surrogate safety measures) are employed. Further, these methods can also be categorized as statistical methods (e.g., accident indicators, accident rates, etc.), probabilistic methods (based on statistical distributions), and vehicle movements-based methods (as per conflicting situations). In the reactive approach, previous years crash data is used to perform crash rate analysis (i.e., temporal trends) and hotspot analysis (i.e., spatial trends), crash frequency analysis, and crash severity analysis. The crash rate analysis is performed to analyze the number of crashes per unit of time. The crash frequency analysis is performed to find the possible number of crashes per section of the road. The crash severity analysis provides the severity of the crashes (fatal, severe injury, minor injury, no injury). Typically, the integrated approach is used, which involves two or more of these methods.

Overview of SPF:

Safety Performance Function (SPF) is a crucial tool for understanding the relationship between various factors and the probability or severity of traffic accidents. It is also known as a crash prediction model, road safety model, or crash frequency analysis. The SPF involves analysing historical data and considering traffic volume, road structure, weather conditions, and other variables (driver characteristics and vehicle attributes) [3]. It creates a systematic approach to identifying high-risk areas and estimating potential accident rates. This process also informs evidence-based decision-making for road safety improvements, enabling authorities to pinpoint specific risk factors contributing to roadway accidents. It facilitates the targeted implementation of effective countermeasures and optimises resource allocation by focusing interventions where they can have the most significant impact. SPF also aid in the analysis of the effects of various road upgrades, such as junction improvements, cross-section enhancements, and road repair and rehabilitation.

This SPF is constructed through statistical or machine learning-based models, sometimes incorporating crash modification factors (CMFs). In statistical or machine learning-based models, jurisdiction-specific SPFs are formulated by establishing relationships between crashes and influencing variables at crash sites [4]. Alternatively, CMFs predict crash frequency for road segments under base conditions, with adjustments made for segments not meeting these conditions based on crash modification factors for each influencing variable [4]. Further, the Calibration factors help to transfer findings between regions, with CMFs representing the ratio of average crashes for the base and non-base conditions and calibration factors reflecting the ratio of observed to predicted crash values. Comparatively, jurisdiction-specific SPFs outperform CMFs-SPFs due to the latter's aggregation of crash data, potentially leading to prediction errors [5]. This study focuses on jurisdiction-specific SPFs, which demand substantial site-specific crash data, making them resource-intensive.

The SPFs, in conjunction with condition and collision diagrams, provide a visual overview of accident sites, collision types, and influential factors. These diagrams are superimposed over road layouts and geometric designs, allowing engineers to precisely detect design flaws that may lead to accidents. This visual aid enhances the understanding of accident patterns, informing the development of effective design solutions and safety enhancements.

Objective of the study: None of the past studies has attempted to look at the data sources and availability of geometric data to develop a Safety Performance Function, which is very critical. There is a need to understand the influence of different factors (e.g., road characteristics, weather, time of the crash, etc.) on crashes and the availability of data sources. Of them, road geometry plays a significant role. Therefore, the objective of this paper is to review the different geometric variables, which are needed for developing an efficient SPF for highways or mid-block sections, the data sources and their availability.

2 Geometric variables

The explanatory variables can be classified into geometry road furniture and facilities (like Signs, markings, lighting facilities, parking, bus/ truck lay byes, pedestrian facilities, access control, toll plaza, and noise barriers), traffic (volume, speed, and vehicle composition), performance (such as pavement type, friction, roughness, rutting depth), and others (includes Driver Characteristics, Vehicular Characteristics, Weather, land use, and roadside hazard rating). Geometric variables are of the highest significance, as poor geometric conditions might create an unsafe scenario for road users.

The following sections discuss each of the geometric variables with respect to past studies. A summary of the geometric variables and their classification categories are listed in Table 1.

2.1 Lane

Most of the past studies have included the number of lanes in the development of SPF [6–17]. A few studies have used the lane width [4, 10, 11, 15, 16, 18–20] and pavement width [14].

A common inference is that an increase in the number of lanes can increase the chances of crashes. That is, with an increase in space availability, the drivers may speed up and/ or become lesser attentive. Further, more space availability increases the opportunities and thus, a higher number of lane change events [17]. In comparison to two lanes, three-lane, four-lane, and five-lane dual carriageways had 56%, 171%, and 161% higher crashes, respectively. In the case of single-vehicle crashes, more fatal and injury crashes are observed with an increased number of lanes [8], which confirms the distraction or higher speeds. In one of the studies, the number of lanes is reported as insignificant for multi-vehicle crashes [8, 21]. Interestingly, increasing the number of lanes increases crash frequency for 78.16% of highway segments but reduces it for the rest. A plausible reason for the latter is that increased available space allows drivers to react to unexpected events by deviating from the path to avoid a collision.

Similar to the effect of an increase in the number of lanes, a few studies reported that an increase in the lane width increases road crashes [10, 18, 19]. Wang et al and Wen et al [18, 19] reported that most of them are single-vehicle crashes. Moreover, few other studies contradict it. For instance, Agbelie [11] reported that an increase in the land width increases the injury-type crashes for 78.16% of the road segments. Similarly, if lane width is less than 3.5m for principal arterial roads, accident frequencies are observed to decrease in Eastern Washington [16]. If the lane width is less than 12ft (approximately 3.65m)

Table 1: Geometric variables

variable	categories	
	quantitative	qualitative
lane	number of lanes, lane width, pavement width	
shoulder	shoulder width	type of shoulder, shoulder condition, location of the shoulder (inside/ outside)
median	median width	presence of median, median type
terrain	longitudinal grade, slope change rate, number of slope changes per vertical profile	presence of gradient, terrain type
curvature	degree of curvature, curvature, vertical curve, length of the curve, horizontal Sinuosity, curve density	presence of curve
visibility	sight distance, horizontal clearance	
bridges/ flyovers/ underpass/ tunnel	% of segment with a structure	presence of structure
superelevation	superelevation value, variance between designed and existing superelevation	

reduces minor crashes for two-lane rural roads [22]. In the case of state highways, a lane width greater than 10ft (approx 3m) reduces the crash frequency of property damage crashes [20].

A few studies have tried to see the significance of lane width, depending on different road types. For instance, the lane width has no or less significance on paved country roads [20] and on two-lane dual carriageway road segments [4]. Similarly, in a study by Das et al [15] on Ohio and Washington for all crash types, lane width is not significant for rural multi-lane highways for both states and two-lane rural roads of Ohio. But an increase in lane width is reducing the number of crashes on two-lane rural roads of Washington. This variation between states is probably because of higher traffic on rural two-lane roads of Washington state than that of Ohio. So, on increasing lane width, there is free space available for maneuvers to avoid crash conditions.

Instead of the number of lanes and lane width, Weng et al [14] considered pavement width and reported that 1% increase in pavement width is reducing crash frequency by 39% and 15.9% for traffic hazards-based and non-traffic hazards-based models, respectively.

2.2 Shoulder

Most of the studies included shoulder details for developing SPF in the form of shoulder width (outer side shoulder or inside shoulder with respect to road) [1, 4–10, 13, 15, 16, 18, 21–31], shoulder type [10, 20, 27, 31, 32], presence of shoulder[33, 34] and shoulder conditions (good, poor) [35].

Typically, the broader shoulders boost road capacity and provide a place for broken vehicles to prevent crashes [23]. The most usual inference obtained from the aforementioned studies is that an increase in shoulder width (outer or inner) reduces the number of crashes. The wider shoulder provides drivers more room to correct their deviation from path/ maneuver [24]. Thus, the narrow shoulder (outer and inner) widths of less than 1.5m (5ft) are increasing accidents for principal arterial roads [16]. Moreover, a decrease in collisions is observed for shoulder width (outer and inner) smaller than 10 feet for highways [27].

Islam et al [8] reported that outer shoulder widths greater than 10ft are safer with respect to single-vehicle crashes of fatal and injury types. Similarly, it was discovered that an increase in the shoulder width reduced the average yearly crash frequency by 0.067 units [26]. On the contrary, increasing shoulder width increases the number of crashes [10] for two-lane rural roads and highways, respectively. Similarly, Malyshkina and Mannering [6], outer shoulders wider than 5 feet enhance the risk of crashes on interstate highways. This is possibly due to the speeding, given more space is available.

Similar to outer shoulder widths, increasing inside shoulder widths also reduces crashes [10, 24]. Especially of run-off road crash type [18]. Anastasopoulos and Mannening [28] found lower crashes for inner shoulder width of 1.5m or higher in 89.49% of rural interstate highway segments and higher crashes for the rest. A similar trend is observed for inner shoulder, for different levels of crash severity [29]. Das et al [36] used rule-based machine learning algorithms to verify the relationship of runoff road crashes with shoulder width formulating rules based on segment lengths, AADT, and shoulder width. It was observed that shoulder width is exhibiting negative relation with the number of crashes in most of the rules for a specific range of AADT and shoulder width. According to Garach et al [34], the shoulder has a significant impact on crashes if AADT is less than 4000 vehicles per day, regardless of whether the shoulder is paved or not, and a 10% increase in the percentage of the road having shoulder is reducing crashes by 5%. However, the shoulder is not significant on highways with an AADT of more than 4000 vehicles per day.

According to Megat-Usamah [9], increasing the left (inner) shoulder width reduces collisions in Minnesota, Ohio, and Washington. Right (outside) shoulder width reduces collisions in Ohio and Washington while increasing crashes in Minnesota. This opposing trend might be the impact of the different number of lanes on each state's highways. As roads with higher lane widths may not need much wider shoulders. Interestingly, Venkataraman et al [25] tried to study the number of crashes for three different categories (3-4 feet, 5-9 feet, and >10 feet) of shoulder widths for the left (inner) and right (outer) shoulders. The comparisons were made with respect to the shoulder width of 2 feet. It was found that inner shoulder width is contributing more to reducing the number of crashes due to a higher provision of overtaking opportunities. For fatalities in crashes, interstate roads are quite sensitive to changes in the inner shoulder width, but they are less susceptible to changes in the outside shoulder width [23]. This is because inner lanes are high-speed lanes so, space availability at the inner side provides room for correcting drivers' mistakes.

A few studies also reported that shoulder width is not significant in influencing crashes [1, 7, 30, 31] despite their location [21]. Vayalamkuzh and Amirthalingam [1] stated that this could also be due to no significant variability in the data. The shoulder condition came out to be insignificant for two-lane highways even under heterogeneous traffic conditions [35].

The paved shoulders are reducing collision for all crash types with a higher importance in fatal injury crashes for state highways. But for paved country roads, it is showing contradicting results which may be due to higher traffic volume on paved country roads having paved shoulders than that of roads with unpaved shoulders [20]. The asphalt-type (paved) shoulders are increasing collisions [32] even if the shoulders are less than 10 feet wide [27]. Further, the footpath type shoulder increases vehicular crashes [10]. This may be due to the unavailability of space for the driver to recover without collision if deviated from the path.

2.3 Median

Typically for SPF functions, median is included in the form of median width [10–12, 23, 26, 29–31, 37], median type [8, 28, 38–42], and presence of median [33]. The most common inference is that increasing median width decreases crashes. This is due to the median's ability to separate traffic according to the flow direction, give a recovery area for negligent drivers, a refuge place to avoid impending hazards, and allow for emergency stops [23]. An increase in median width reduces fatal injury crashes and at other severity levels, it doesn't show much significance [37]. In accordance with a study on multi-lane highways [29], a minimum median width of 45ft and 75ft is required to reduce injury crashes and non-injury crashes respectively. Agbelie [11] revealed that a unit increase in median width reduces crashes by 0.003 units. Furthermore, Chen et al [23] reported that median widening by 1% reduces crashes by 0.5021% casualty accidents and 0.694% non-casualty accidents on interstate highways, by 0.1127% casualty accidents for US roads and has no impact on the state roads. Thus, the impact of the median is higher in the case of higher road class which is due to higher traffic flow at higher speeds. The insignificance in state roads is possibly due to the lesser variability of data with different median widths compared to interstate roads and US roads. A wider median helps to reduce overall crashes for urban and rural segments but boosts rollover crashes in rural areas [30].

According to Pan et al [12], an increase in median width reduces crashes in Ontario State and urban multi-lane of Washington. But it increases the crashes on two-lane two-way rural roads in Washington. This is due to the restricted or unavailability of space for overtaking operations.

Similar to the median width, the presence of media also reduces crashes [33, 43], especially of barrier type [28, 38, 39]. Such barriers stop or restrict out-of-control vehicles or hazards from entering the opposite lanes of traffic. Dong et al and Champahom et al [10, 40] stated that non-traversable medians increase safety more than the traversable median. Median type as paved or vegetation reduces crashes but raised median increases crashes of runoff road type. This happens when a vehicle at a higher speed hits the raised median, and loses control; mountable medians provides an opportunity to recover [41]. The Jersey-type median is increasing multiple vehicle crashes of higher severity [42]. According to Islam et al [8], median type is insignificant for single-vehicle crashes.

2.4 Terrain

The studies included vertical grade in the terms of longitudinal grade [4, 11, 13, 18, 23, 34, 44–51], presence of gradient [10, 38], slope change rate [2, 52], number of slope changes per vertical profile [7] and terrain type (plain, rolling and mountainous) [40, 53] in developing SPF.

The most usual finding is that increase in slope increases crashes. This is because of the larger variation of operating speed and this impact is high for heavy vehicles i.e., causing non-homogeneity of traffic flow conditions [54]. As per a study done for different highways by Chen et al [23], higher-class roads are more sensitive than lower-class roads for a given crash severity and casualty accidents are more sensitive than non-casualty accidents for a given road type in increasing crashes with vertical grade. This is because higher-class roads are to be built with high standards and for high speeds. According to Chang and Chen [45] for a freeway, if the grades are greater than 3% or less than -4% the crashes increase and if the grade is between 0 and 1 percentage it reduces crashes. This reduction in crashes at lower grades may be because of the good control of the vehicle and speed by drivers at these grades. A study reported an inconsistent behavior of crashes i.e., vertical grade is greater than 5% is increasing crashes in 58.64% of road sections and decreasing crashes for 41.54% of road sections of highways. A unit increase in the vertical grade above 5% increases crash frequency by 0.121 units [11]. This can be due to the alertness of drivers on slopes and controlled speed. Furthermore, Garber and Ehrhart [46] proved that grade less than 4% does not influence crashes. The influence of longitudinal grade on increasing crashes is very low [34] may be because of having majority of the roads with lesser grade values.

Furthermore, collision frequency on horizontal curves combined with sag vertical curves is generally greater than that for horizontal curves combined with crest vertical curves [55]. It can be due to the downgrade which is causing an increase in speeds combined with centrifugal forces leading to less control over vehicles and increasing crashes.

In few studies, grade values came out to be insignificant [4, 13, 18, 47–49] for a value of less than 6% [54]. This may be due to longitudinal grade correlated with other variables like speed[49] or no much variance in the grade value of the study dataset.

Like in longitudinal grade values, the presence of gradient also increases crashes[10]. Especially for grades sharper than 6% [54] . Intriguingly, Anastasopoulos et al [38], found that vertical grade presence increased crashes in 61.58% of cases only. The reduction in 38.42% cases may be having lower slope values as lower slope values are found to reduce crashes as per previous studies.

An increase in slope change rate increases crashes but the impact is minor [2]. The influence of slope change rate on crashes is also influenced by the section considered for the study. That is, slope change rate reduces crashes when the road section consisting two tangents and two curves are considered as a section and is increasing when fixed length of the road section. It is insignificant when the sections are formulated based on curves or homogeneous section [52]. Besides, vertical sinuosity (algebraic difference of slope by length i.e., slope change rate) increases crashes increases [7].

The longitudinal slope in terms of the number of slope changes per vertical profile along the roadway is insignificant as most of the data considered for study has no change in slope per segment [6].

The length of longitudinal slope corresponding to minimum grade reduces crashes [50, 51]. The variation (elasticity effect) by severity type is high at night time than day time. Besides, daytime has less severe crashes and more minor crashes and night time has less minor crashes and more injury crashes [51]. This might be because of less traffic and visibility issues at night.

Moreover, the terrain type as mountainous increases crashes [40]. The plain terrains are safer than rolling terrain for two-lane NH and insignificant for 6-lane and 8-lane National Highways as the study is in 6-lane and 8-lane NH are at mostly of plane terrain [53].

2.5 Curvature

Most of the studies used horizontal and vertical curve information in the development of SPF in the form of the presence of curve [16, 20, 26, 30, 38, 53], degree of the curvature [1, 6, 12, 13, 17, 18, 28, 35, 38, 48, 54–57], curvature [1, 2, 21, 47, 49, 58], length of curve [16, 25, 30, 34, 50, 51], horizontal sinuosity [7] and curve density [22, 57].

The presence of a curve in a segment increases the chances of crashes by 0.254 units for highways [26]. Similarly, the addition of a horizontal curve to a segment increases the possibility of crash occurrences by 4.44% [25]. The poor design of curves (improper introduction of transition curves, continuous sharp curves, etc.) or inadequate sight distanceare possible reasons for an increase in total crashes. According to Dutta and Fontaine [30], the presence of horizontal curves is increasing crashes possibility for 2-lane bi-directional rural freeway segments. But insignificant in the case of 3-lane bi-directional urban freeway segments of Virginian because the curves mostly present are gentle curves (almost resembling tangent). Similarly, the presence of a horizontal curve came out to be insignificant for two/ four/ six-lanes highways [53]. On the contrary, few studies reported that the presence of a horizontal curve reduces the chances of crashes [16, 20, 38]. This is because the change in roadway geometry increases driver alertness (highway hypnosis) and drivers are more careful in dangerous situations, whereas they are careless (risk compensation) in other situations [38].

From the referred studies that used the degree of curve, the usual inference is that an increase in the degree of the horizontal curve (i.e., sharp curve) increases crashes [1, 2, 13, 48, 54–56] because of the increase in centrifugal force on vehicles negotiating the curves. In Contrast, some studies reported that an increase in the degree of the horizontal curve (i.e., sharp curves) reduces the probability of crashes [6, 12]. Possibly, because of reduced speeds and increased alertness of the driver. Furthermore, a unit increase in the degree of the horizontal curve is decreasing crash probability by 0.142 units [28]. According to Wen et al [18], if the degree of the horizontal curve increases greater than >2 degrees, it increases runoff road crashes. Moreover, if the degree of the curve is >3, an increase in the runoff road crash becomes stable. Similarly, if the degree of the curve is greater than 6 degrees, the possibility of crashes on freeways decreases [17]. On interstate highways, it was found that an increase in the degree of curve reduces crashes in 90.23% of the segments, and in the rest 9.77% of cases it is increasing crashes [38]. In a study degree of the curve came out to be insignificant [49]. This may be because most of the sections have larger radii [47].

Similar to the degree of curve, an increase in curvature also increases crashes, especially of run-off road crash type [58]. According to Vayalamkuzhi and Amirthalingam [1], a drop of 0.1/km horizontal curvature decreases the possibility of a crash by 1.40%. Further Chen et al [21] reported that large curvatures increase crashes in 80.9% of cases and reduce crashes in 19.1% of cases reflecting complex driving behavior while traversing curves or because of proper design of curves like sufficient super elevation, and tangent length. Looking for different vehicle types, the risk of a crash is high in two-wheeler compared to that of four-wheelers due to irregular driving habits of two-wheelers causing a higher danger of rider losing control of the vehicle on horizontal curves [2].

Further, an increase in the length of the front section curve also increased casualty crashes [50] due to the increased overall length of continuous curves. Furthermore, Venkataraman et al [25] reported that 1m reduction on the shortest length of the horizontal curve in the segment, the probability of crashes reduces by 16% for interstates of Washington. According to Zhang et al [51], increasing the radius and length of the present section increases the likelihood of light-injury collisions and night-time crashes. With every 1% increase in front section length, the likelihood of a more serious crash increases by 54%. This may be the result of speeding and fatigued drivers, particularly in the evenings. As per a study considering horizontal curve details in terms of curvature change ratio, if AADT is greater than 4000 vehicles/ day, a 10% increase in the curvature change ratio increases crashes by 24% because increased interactions between surrounding vehicles at curves increases the chances of crashes and if AADT less than 4000 veh/ day curvature change ratio is insignificant [34]. In contrast, a study reported that an increase in the length of the curve reduces crashes, especially for rural roadway segments; the same is insignificant for urban roadway segments [30].

Horizontal sinuosity is defined as the ratio of horizontal curvature per unit length. An increase in horizontal sinuosity reduces the probability of crashes [7]. This needs further investigation as an increase in sinuosity might increase the probability of crashes due to a higher rate of change of curvature. Similarly, curve density, the number of curves per unit length, increases crashes with increased curve density [57]. The rise in accidents is because an increase in curve density subsequently contributes to challenges encountered by drivers in maintaining their lane position [22].

Additionally, tangent lengths greater than 0.8km and 0.4km increase the crashes for western and eastern Washington, respectively [16]. This might probably be because of speeding on the long tangent and being unable to get opportunity to prepare for the next curve.

The studies that used the vertical curves in SPF development considered the number of vertical curves per mile [25, 26, 28], vertical curve length [28], and vertical curvature [1]. According to Venkataraman et al [25], an increase in the number of vertical curves increases crashes. An increase in the number of vertical curves per mile causes an increase in crashes for 84.13% of road segments, and a unit increase in vertical curves increases crashes by 0.06 units [26]. This may be because the drivers are less familiar with the road alignment ahead. Whereas in 15.87% of segments, crashes decreased possibly due to a reduction in the drowsiness of drivers due to changes in geometry [26]. It can also be due to proper signs and markings provided in those sections giving the information about the road ahead. In contrast, the number of vertical curves per mile reduces crashes, and for every vertical curve added, crashes reduce by 0.199 [28].

Moreover, Anastasopoulos and Mannering [28] used the ratio of vertical curve length to the segment length in SPF development and found that a unit increase in the ratio, increases crashes by 0.142 units. This happens because the sharp vertical curve increases the difficulty of driving (especially for heavy vehicles), and poor sight distance.

2.6 Visibility

A few studies used visibility in developing SPF in terms of sight distance [47, 48, 50, 51, 54]. The usual inference is that an increase in sight distance reduces the possibility of crashes, because of better facilitation for drivers to control the maneuver accurately to avoid any possible conflict [50]. Conversely, excellent driving conditions with higher sight distances will induce drivers to exceed the speed limit[54]. Similarly, an increase in the sight distance and horizontal clearance for trucks increases crashes; however the intensity is low since trucks had a larger distance to react properly to an incoming crash [51]. A study reported sight distance as insignificant (at 5% level) in four-lane divided motorways [48]. This may be due to the required amount of sight distance in most of the sections.

2.7 Bridges/ Flyover/ Underpass/ Tunnel

A few studies used the presence of a bridge in SPF development. The presence of bridges is reducing the crashes [6, 28] due to good design standards. In contrast, the presence of a bridge is increasing the probability of crashes by around 0.3-0.43 units possibly due to defects in the design of the bridge or visibility issues [35].

Similar to the presence of bridges, the percentage of the segment with a tunnel reduces crashes when the section considered is a homogeneous section with respect to AADT and curvature. Whereas, it is increasing crashes if the roadway segment is fixed, based on curves, and two curves two tangents type division [52]. In the case of homogeneous sections, the section lengths can be smaller so, the reduction in crashes is due to the alertness of drivers due to changed cross sections. But in other road division methods, the segment length is probably high, increasing crashes due to the monotony of drivers. Furthermore, the tunnels may have traffic hazards, and poor visibility; there are chances of diminished driver concentration due to longer tunnels, which may increase the chances of crashes [14].

2.8 Superelevation

very few studies used superelevation and differences between existing and designed superelevation values in SPF development. It was observed that an increased superelevation reduces collisions southbound because it compensates for the influence of centrifugal force at sharp curves. Whereas, increasing collisions of northbound can be due to insufficient superelevation [1]. Further, superelevation is insignificant for four-lane divide highways [49]. Moreover, an increase in the variance between the existing and designed superelevation increases the crashes in Kerman province of Iran [13].

3 Data Sourcing

3.1 Variables in the model development

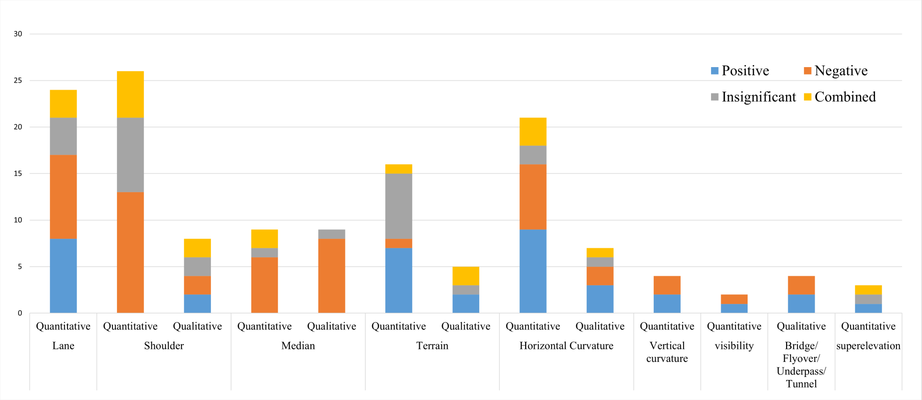

Fig. 1: Variables literature count and their relationship with crashes for Highways.

From the aforementioned discussion, it can be stated that the data required to develop a Safety Performance Function is high. However, various studies have used different numbers of variables, mostly as to the availability of the data. Figure 1[1] depicts the frequency of studies undertaken on various highway sections, which included an examination of numerous geometrical variables either qualitatively or quantitatively (c.f., Table 1). Similarly, Figure 2 depicts the frequency

---

[1] The figure is prepared using 89 studies. However, all of these studies are not cited in this work due to the limitation of the number of pages. All studies with important or unique inferences are included in this work.

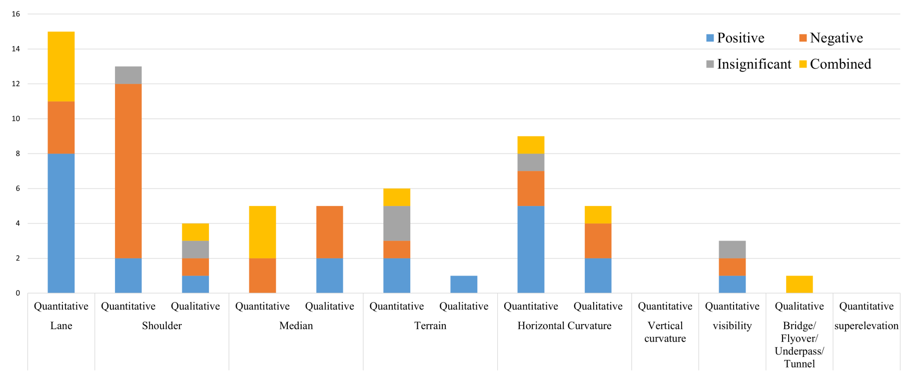

Fig. 2: Variables literature count and their relation with crashes for other road classes or multiple road classes.

of studies undertaken on various road sections other than highways, including freeways, multiple-class roads, rural roads and others. These studies are also assessed for examination of numerous geometrical variables either qualitatively or quantitatively. It was observed that Lane, shoulder and horizontal curvature information is used in most studies. The variables which are difficult to obtain (e.g., sight distance, superelevation) are used in a few studies only. Moreover, many studies have used lane, shoulder, terrain and horizontal curvature measures quantitatively. Figure 1 and Figure 2 also exhibit a number of studies that shows different trends (positive, negative) for each of the geometrical variables. The positive trend means the number of crashes and the variable move together in a direction (e.g., increasing lanes increases the number of crashes). Moreover, some of the studies find one or more variables insignificant. This further reveal the distinct pattern for highways and other roads. The highways show varying relationships between literature count on lane width and crashes. In contrast, other roads exhibit a pronounced dominance of positive relationships. Quantitative measures of shoulder width indicate predominantly negative relationships on highways, with some insignificant and combined relationships, while other roads exhibit a mix of negative and a few positive relationships. Regarding quantitative median measures, highways mostly display negative relationships with some significant and combined associations, while other roads show mixed negative and combined relationships. Additionally, the qualitative assessment of median measures indicates that highways have a prevalent negative crash association. In contrast, other roads display a varied mix of positive and negative relationships regarding medians' impact on crash risk. The variation in these patterns is likely due to differences in their design standards, traffic volume, rural and urban road settings, data quality, and geographic characteristics.

Further, the study area may also affect the variables included. For instance, the presence of structures (bridges, flyovers, tunnels, underpasses) cannot be considered in their absence.

3.2 Source of data

Typically, the crash data is collected from various crash databases, managed by the state of national-level agencies (e.g., Crash Analysis Reporting System by Florida DoT), records from the police station (e.g., First Information Reports), and sometimes, directly in the field. The data collected in the field is mainly the missing information for the reported crashes.

For the geometrical variables, the different measurement forms and data collection techniques, used in the past studies, are listed in Table 2. Lanes are considered in most of the studies (see Figure 1 and Figure 2), in the form of the number of lanes, lane width, and pavement width.

In past, such information is extracted using Google street view maps, Google Earth images, digital terrain models with satellite images, and Network Survey Vehicle (NSV). The former three are geospatial techniques, whereas the latter is a field data collection approach, which collects data for many other variables. Similar techniques, as for lanes, are used in shoulder data extraction for shoulder width and type. But the shoulder condition can be extracted by field observations, video recordings, and NSV.

The Median data is extracted from Google Earth, NSV, field survey, and 3D areal survey. Sometimes, this information is available in the crash report [10]. Extracting terrain data is difficult to get by 2D maps like Google Earth maps or satellites. Therefore, other sources are drawings from design and construction stages, NSV, and 3D areal surveys.

Curvature data is also collected using field surveys, drawings, NSV, 3D areal surveys, digital terrain models, and field Observations. Visibility is obtained from drawings. But this data might not be correct due to a change in visibility at the site due to overgrown trees or unauthorized billboards or other roadside developments.

Bridges/ Flyovers/ Underpass/ Tunnels used in the form of the presence of structure and % of segment with the structure. This information are evaluated for their presence by field observation followed by video recordings. Superelevation is also not extractable by satellite images. So, extracted by field surveys, NSV, and digital terrain models (also called Digital Elevation Models).

OpenStreetMap is an open-source data site. The road network data with details on the number of lanes, one-way or two-way, and the presence of structures are available. Since it is a volunteered map service, it may not be updated for the recent updates. The street view map [59–62], Google Earth(or satellite images) [4, 20, 37, 53, 63–65], etc., are easy to employ, and low cost, but it is laborious and time-consuming. The curvature, slope, sight distance, vertical curvature, slope change rate, etc. are difficult to measure using such data.

The Digital Terrain Model (DTM) along with satellite image-based data collection accuracy is completely dependent on the availability of high-resolution DTM and satellite imagery, and the methodology used to combine them[65]. The above sources can provide good data for smaller as well as for larger regions with an increasing effort from smaller to larger regions.

Field surveys [13, 43, 53]are employed using advanced surveying equipment (e.g, total station) along with GPS might help extract the road's geometric fea-

Table 2: Different geometric variables form and the used data collection methods

Geometric variable	Measurement forms	Data collection method
Lane	Number of lanes	Street-view map
	Lane width	Google Earth, digital terrain models + satellite images, field survey
	Pavement width	Google Earth, Network Survey Vehicle
Shoulder	Presence of shoulder	
	Shoulder width	Google Earth, Network Survey Vehicle, digital terrain models + satellite images, field survey
	Shoulder type	Network Survey Vehicle
	Shoulder location	
	Shoulder condition	field observation + video recorder, Network Survey Vehicle
Median	Presence of median	field survey
	Median width	Google Earth maps
	Median Type	Highway crash reports
Terrain	Presence of gradient	Network Survey Vehicle
	Longitudinal grade	Network Survey Vehicle, 3D aerial survey, built drawings
	Slope Change Rate	built drawings
	Number of Slope Changes per vertical profile	
	Terrain type	strip plan, field visit
Curvature	Presence of Curve	field observation + vedio recorder, strip plan, drawings
	Degree of Curvature	built drawings
	Curvature	Network Survey Vehicle, 3D aerial survey, built drawings, digital terrain models + satellite images
	Vertical Curvature	3D aerial survey, Network Survey Vehicle
	Length of Curve	3D aerial survey, built drawings
	Horizontal/ Vertical Sinuosity	Field Survey
Visibility	Sight Distance and Horizontal Clearance	Drawings, 3D areal survey
Bridge/ Flyover/ Underpass/ Tunnel	Presence of structure	field observation + video recorder
	% of segment with the structure	
Superelevation	Value	Field surveys, Network Survey Vehicles, digital terrain models + satellite images
	Variance in design and existing superelevation	

tures. This needs expertise for handling equipment and is time-consuming for long stretches. The 3D areal survey [48] avails for extracting possibly all the geometric variables. But is costly as it includes using drones or other automated flight systems. The NSV [1, 49, 56, 66] gives accurate details of all road geometric elements and their current conditions with coordinates. However, the data collection from NSV is resource-intensive.

Drawings [2, 53, 67] of the design and construction stage are the cheapest, fastest method of data collection for smaller as well as larger regions. However, this does not incorporate changes in road geometric conditions at a later stage or during the crashes. Video recorder [61] alone can be used in extracting the lane, shoulder, median, visibility, and presence of structure details. The accuracy of the data extraction is not guaranteed as it depends on the angle of mounting the camera, perspective fixation, algorithm, and the speed of the vehicle on which it is mounted.

A majority of these methods don't give the values (or condition) at the time of the crash as they might deteriorate with time or improvise their actual condition because of road maintenance work, as the study period usually varies between 3-5 years. For example, effective lane width might be reduced due to roadside hazards (like parking, retail shops, etc.), poor unpaved shoulder conditions due to severe rain, etc. If their condition at the time of the crash is collected along with accident details in crash reports might be helpful to analyze the conditions. Historical satellite image availability can also help in this aspect, but getting all roadway facilities like elevation and terrain cannot be guaranteed from satellite images. In order to overcome this issue, developed countries like the USA have the Department of Transportation and HSIS (Highways Safety Information System) maintain such data.

4 Research Gap and future directions

The road geometric variables exhibit dual nature (as shown inFigure 1 and Figure 2), i.e., increase or decrease for the same value range based on a set of conditions. For example, on freeways, a shoulder width greater than 10 m reduces crashes, but considering it along with the speed limit, it increases crashes. Further, the findings are heterogeneous with respect to crash type. Clearly, this emphasises the need for the inclusion of all the possible interrelationships between variables (also other than geometric variables) in the development of an SPF model for different crash types. This even helps in decisions on combined and cost-effective solutions for reducing crashes. As shown in Figure 1 and Figure 2, it is possible that a range of a variable reduces the crashes, and beyond this, the variable increases the crashes or becomes insignificant. Clearly, very few studies have found a combined or heterogeneous effect (positive and negative, see Figure 1 and Figure 2), depending on the different characteristics and classification of the geometric variables in the model development. Therefore, future research should consider the multiple classifications of the geometric variables, such that the countermeasures can be tailored specifically to the inferences for given classes of geometric variables.

However, only a few studies possibly included all these factors due to the lack of data or complexity in the analysis. Especially sight distance, bridge/ tunnel length, and superelevation are rarely utilized in the analysis. Interestingly, if the shoulder is used as a qualitative measurement, a significant number of studies are reporting an increase in crashes, which contradicts the general intuition. Similarly, if the median is used as a quantitative measure, none of the studies found a positive or insignificant trend. Moreover, if the terrain is used as a quantitative measure, none of the studies found that an increase in grade/ slope change rate, etc. (c.f., Table 1) will reduce the crashes. Though more studies are using quantitative measures of the geometric variables, the literature lacks in clarity on the preference among qualitative and quantitative measures for different variables. Looking at the number of studies (Figure 1 and Figure 2), in future research, shoulder, terrain, and horizontal curvature can be included qualitatively. However, the decision about the qualitative and quantitive nature of the geometric variables may depend on the availability of the existing data, resources (funds and time) for data collection, etc. Some of the variables are used quantitatively (e.g., superelevation, vertical curvature, sight distance), which are difficult to collect in the field, especially for a larger region. Thus, an attempt can be made to include these variables as qualitative measures to study the impact.

When specific road geometric variables are studied, the practical conditions might differ from the data in the drawings. For instance, the effective lane/ pavement/ shoulder width may be lesser due to encroachments, on-street illegal parking, street hawkers, etc. As a result, considering effective lane width is more beneficial than using actual lane width. There is also a need for research on shoulder conditions since damaged shoulders may increase the likelihood of a crash due to a loss of functionality. Similarly, a damaged median could constitute an illegal entrance or a potentially dangerous roadway feature; thus, such factors can be included qualitatively for the median. Given the topography, the gradient and its length may influence collisions, i.e., steeper gradients for a longer length may increase crashes. As a result, it is important to investigate the impact of both gradient and its length on collisions. There is also a need for further research into visibility, bridge/ tunnel/ flyover/ underpass, and superelevation. It is also essential that a proper understanding of the influence of road geometry variables in diverse road environments and traffic conditions is required.

The majority of research employs roads from the same feature class and for a certain stretch length having similar nature of traffic and road geometric conditions might lead to a misinterpretation of the whole nature of a parameter. As a result, the use of different road types for analysis is beneficial. Further, the study of crashes on urban and rural roads of lesser importance and on multiple road classes are needed.

The majority of the studies are from developed economies and the findings may not be transferable to developing economies; thus, more such works are required for the latter countries. The various aspects observed in developing are presence of only the outer shoulder, in addition to serving emergency parking and storage for broken-down vehicles, the shoulder may also serves as a pedestrian walkway causing an increase in pedestrian-vehicle conflict probability mainly on rural highways. The raised median increases the fatality of crashes on high-speed roads [41, 42] whereas, mountable or paved, or vegetated medians increase unauthorized median openings by motorized two-wheelers and tractors. In hilly regions, it is unavoidable to have higher gradients with hairpin beds due to high excavation costs while higher slopes create difficulty for slow-moving and heavy vehicles. Hence, this necessitates immediate and comprehensive research in this area.

The availability of data is observed to have a great influence on variables selected for data collection. Therefore, there is a need for cost-effective data sources with road geometric variables and their condition with less time for extracting road geometric variables data and a sufficient level of accuracy. Incorporating road condition details at the time of the crash in crash reports might be helpful in understanding the scenario at road geometry at the time of the crash. If the geometric data at the time of a crash is unavailable, other sources (e.g., drawings, historical satellite images) can be used to include them in the model development. Given the ease in availability of high-resolution satellite images and street view data, many of the geometrical attributes may be collected using such advanced sources. Clearly, this may involve more efforts to extract the data and can increase the time to develop the model.

Given the importance of the data related to geometric variables in understanding the crash frequency, and severity, efforts should be made to make the data accessible. A hierarchical institutional structure at the country and state levels may help in data collection, data storage, data management, and distribution. The hierarchy will help in the collection of local road inventory data at the time of the crash.

5 Conclusion

A crash is a rare and complex phenomenon, and the SPF is developed to relate crashes with crash causation. This paper provides a review of Geometric factors influencing crashes and sources of data collection. The road geometry variables that influence crashes, including lane, shoulder, median, terrain, sight distance, curvature, superelevation, and bridge/ tunnel/ underpass/ overpass. Most of the studies use something other than sight distance, structure information, and superelevation due to the ease in availability of such data. Further, all the factors are rarely used due to the complexity of collecting such massive data and analysis complexities. Incorporating all the road geometry factors into SPF may improve knowledge of crashes and aid in developing corrective measures. It is also observed that heterogeneity in the geometric variables for a road type is higher in developing economies, which may influence crash frequency and severity, and thus, more work is required in developing countries. The other aspects needed for further investigation are the interrelation between variables, the study of crashes in rural and urban roads of lesser importance, and the impact of continuous changes in road geometry and environment. The different data sources, survey techniques for data collection, and various forms of variable information that can be collected are also reviewed. The need for better data sources is observed in developing countries which overcomes by advanced methodologies. Therefore, there is a need to develop a cost-effective and accurate data extraction process with a possible level of automation.

Bibliography

[1] P. Vayalamkuzhi and V. Amirthalingam. Influence of geometric design characteristics on safety under heterogeneous traffic flow. *Journal of Traffic and Transportation Engineering (English Edition)*, 2016.

[2] N. K. ChikkaKrishna, M. Parida, and S. S. Jain. Calibration of safety performance function for crashes on inter-city four lane highways in India. *Cogent Engineering*, 2015.

[3] J. Ambros, C. Jurewicz, S. T., and M. Kieć. An international review of challenges and opportunities in development and use of crash prediction models. *European Transport Research Review*, 2018.

[4] B. Brimley, M. Saito, and G. Schultz. Calibration of highway safety manual safety performance function. *Transportation Research Record*, 2012.

[5] G. Mehta and Y.Lou. Calibration and development of safety performance functions for Alabama. *Transportation Research Record*, 2013.

[6] N. V. Malyshkina and F. L. Mannering. Zero-state markov switching count-data models: An empirical assessment. *Accident Analysis and Prevention*, 2010.

[7] J. O. da Costa, A. P. J. Maria, P. A. A. Pereira, E. F. Freitas, and F. E. C. Soares. Portuguese two-lane highways: modelling crash frequencies for different temporal and spatial aggregation of crash data. *Transport*, 2015.

[8] M. S. Islam, J. N. Ivan, N. E. Lownes, R. A. Ammar, and S. Rajasekaran. Developing safety performance function for freeways by considering interactions between speed limit and geometric variables. *Transportation Research Record*, 2014.

[9] M. U. Megat-Johari and J.Cyr. A comparison between Minnesota, Ohio, and Washington using hsis data transferability of safety performance functions across jurisdictions. *Institute of Transportation Engineers.ITE Journal*, 2018.

[10] T. Champahom, S. Jomnonkwao, C. Banyong, W. Nambulee, A. Karoonsoontawong, and V. Ratanavaraha. Analysis of crash frequency and crash severity in thailand: Hierarchical structure models approach. *Sustainability*, 2021.

[11] Bismark R. D. K. Agbelie. Random-parameters analysis of highway characteristics on crash frequency and injury severity. *Journal of Traffic and Transportation Engineering (English Edition)*, 2016.

[12] G. Pan, L. Fu, and L. Thakali. Development of a global road safety performance function using deep neural networks. *International Journal of Transportation Science and Technology*, 2017.

[13] M. Haghani, R. Jalalkamali, and M. Berangi. Assigning crashes to road segments in developing countries. *Proceedings of the Institution of Civil Engineers - Transport*, 2019.

[14] J. Weng, X. Gan, and J.Chen. A separate analysis of crash frequency for the highways involving traffic hazards and involving no traffic hazards. *Journal of Transportation Safety and Security*, 2021.

[15] S. Das, S. R. Geedipally, and K. Fitzpatrick. Inclusion of speed and weather measures in safety performance functions for rural roadways. *IATSS Research*, 2021.

[16] J. Milton and F. Mannering. The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. *Transportation*, 1998.

[17] Li-Yen Chang. Analysis of freeway accident frequencies: Negative binomial regression versus artificial neural network. *Safety Science*, 2005.

[18] X. Wen, Y. Xie, L. Jiang, Y. Li, and T. Ge. On the interpretability of machine learning methods in crash frequency modeling and crash modification factor development. *Accident Analysis and Prevention*, 2022.

[19] K. Wang, J. N. Ivan, N. Ravishanker, and E. Jackson. Multivariate poisson lognormal modeling of crashes by type and severity on rural two lane highways. *Accident Analysis and Prevention*, 2017.

[20] M. Chakraborty and T. J. Gates. Association between driveway land use and safety performance on rural highways. *Transportation Research Record*, 2020.

[21] F. Chen, S. Chen, and X. Ma. Analysis of hourly crash likelihood using unbalanced panel data mixed logit model and real-time driving environmental big data. *Journal of Safety Research*, 2018.

[22] Asif Mahmud, Vikash V. Gayah, and Rajesh Paleti. Estimation of crash type frequency accounting for misclassification in crash data. *Accident Analysis and Prevention*, 2023.

[23] S. Chen, T. Usman Saeed, M. Alinizzi, S. Lavrenz, and S. Labi. Safety sensitivity to roadway characteristics: A comparison across highway classes. *Accident Analysis and Prevention*, 2019.

[24] Qi Shi, M. Abdel-Aty, and R. Yu. Multi-level bayesian safety analysis with unprocessed automatic vehicle identification data for an urban expressway. *Accident Analysis and Prevention*, 2016.

[25] N. S. Venkataraman, G. F. Ulfarsson, V. Shankar, J. Oh, and M. Park. Model of relationship between interstate crash occurrence and geometrics. *Transportation Research Record*, 2011.

[26] Bismark R. D. K. Agbelie. A comparative empirical analysis of statistical models for evaluating highway segment crash frequency. *Journal of Traffic and Transportation Engineering (English Edition)*, 2016.

[27] R. Krueger, P. Bansal, and P. Buddhavarapu. A new spatial count data model with bayesian additive regression trees for accident hot spot identification. *Accident Analysis and Prevention*, 2020.

[28] P. Ch Anastasopoulos and F. L Mannering. A note on modeling vehicle accident frequencies with random-parameters count models. *Accident Analysis & Prevention*, 2009.

[29] T. U. Saeed, T. Hall, H. Baroud, and M. J. Volovski. Analyzing road crash frequencies with uncorrelated and correlated random-parameters count

models: An empirical assessment of multilane highways. *Analytic Methods in Accident Research*, 2019.

[30] N. Dutta and M. D. Fontaine. Improving freeway segment crash prediction models by including disaggregate speed data from different sources. *Accident Analysis and Prevention*, 2019.

[31] Ahmed Farid, Mohamed Abdel-Aty, and Jaeyoung Lee. A new approach for calibrating safety performance functions. *Accident Analysis and Prevention*, 2018.

[32] P. Buddhavarapu, J. G. Scott, and J. A. Prozzi. Modeling unobserved heterogeneity using finite mixture random parameters for spatially correlated discrete count data. *Transportation Research Part B*, 2016.

[33] J. Q. Li and W. Yu. Enhanced safety performance function for highway segments in Oklahoma. *Journal of Infrastructure Systems*, 2021.

[34] L. Garach, J. de Ona, G. LOpez, and L. Baena. Development of safety performance functions for spanish two-lane rural highways on flat terrain. *Accident Analysis and Prevention*, 2016.

[35] S. M. Sohel Mahmud, L. Ferreira, Md S. Hoque, and A. Tavassoli. Micro-level safety risk assessment model for a two-lane heterogeneous traffic environment in a developing country: a comparative crash probability modeling approach. *Journal of Safety Research*, 2019.

[36] S. Das, X. Sun, and M. Sun. Rule-based safety prediction models for rural two-lane run-off-road crashes. *International Journal of Transportation Science and Technology*, 2021.

[37] K. A. Alkaaf and M. A. Aty. Transferability of safety performance functions: the case of urban four-lane divided roadways in Muscat. *Vehicles*, 2022.

[38] P. Ch. Anastasopoulos, F. L. Mannering, V. N. Shankar, and J. E. Haddock. A study of factors affecting highway accident rates using the random-parameters tobit model. *Accident Analysis and Prevention*, 2012.

[39] A. Das and M. A. A. Aty. A combined frequency-severity approach for the analysis of rear-end crashes on urban arterials. *Safety Science*, 2011.

[40] C. Dong, C. Shao, J. Li, and Z. Xiong. An improved deep learning model for traffic crash prediction. *Journal of Advanced Transportation*, 2018.

[41] M. M. Ali Al-Omari, M. Abdel-Aty, and Q. Cai. Crash analysis and development of safety performance functions for Florida roads in the framework of the context classification system. *Journal of Safety Research*, 2021.

[42] C. N. Kumar, M. Parida, and S.S. Jain. Poisson family regression techniques for prediction of crash counts using Bayesian inference. *Procedia - Social and Behavioral Sciences*, 2013.

[43] Z Sawalha and T Sayed. Traffic accident modeling: some statistical issues. *Canadian Journal of Civil Engineering*, 2006.

[44] A. Abdulhafedh. Crash frequency analysis. *Journal of Transportation Technologies*, 2016.

[45] Li-Yen Chang and Wen-Chieh Chen. Data mining of tree-based models to analyze freeway accident frequency. *Journal of Safety Research*, 2005.

[46] N. J. Garber and A. A. Ehrhart. Effect of speed, flow, and geometric characteristics on crash frequency for two-lane highways. *Transportation Research Record*, 2000.

[47] J. M. Pardillo Mayora, R. B. Manzo, and A. C. Orive. Refinement of accident prediction models for spanish national network. *Transportation Research Record*, 2006.

[48] C. Caliendo, M. Guida, and A. Parisi. A crash-prediction model for multilane roads. *Accident Analysis and Prevention*, 2007.

[49] P. Vayalamkuzhi and V. Amirthalingam. Development of comprehensive crash models for four-lane divided highways in heterogeneous traffic condition. *Transportation Research Procedia*, 2016.

[50] C. Wang, F. Chen, J. Cheng, Wu Bo, P. Zhang, M. Hou, and F. Xiao. Random-parameter multivariate negative binomial regression for modeling impacts of contributing factors on the crash frequency by crash types. *Discrete Dynamics in Nature and Society*, 2020.

[51] P. Zhang, C. Wang, F. Chen, S. Cui, J. Cheng, and W. Bo. A random-parameter negative binomial model for assessing freeway crash frequency by injury severity: daytime versus nighttime. *Sustainability*, 2022.

[52] S. Cafiso, C. D'Agostino, and B. Persaud. Investigating the influence of segmentation in estimating safety performance functions for roadway sections. *Journal of Traffic and Transportation Engineering (English Edition)*, 2018.

[53] H. M. Naqvi and G. Tiwari. Safety performance functions for fatal crashes on national highways in India. *Transportation Research Procedia*, 2020.

[54] E. M Choueiri, R. Lamm, J. H Kloeckner, and T. Mailaender. Safety aspects of individual design elements and their interactions on two-lane highways: international perspective. *Transportation Research Record*, 1994.

[55] Said M. Easa and Qing Chong You. Collision prediction models for three-dimensional two-lane highways. *Transportation Research Record*, 2009.

[56] A.Arun, E. Madhu, and S. Velmurugan. Road crash frequency prediction for Indian national highways using soft-computing tools. In *Recent Advances in Traffic Engineering*, 2015.

[57] Vikash V. Gayah and Eric T. Donnell. Estimating safety performance functions for two-lane rural roads using an alternative functional form for traffic volume. *Accident Analysis and Prevention*, 2021.

[58] J. W. H. van Petegem and Fred Wegman. Analyzing road design risk factors for run-off-road crashes in the netherlands with crash prediction models. *Journal of Safety Research*, 2014.

[59] P. Barbosa Silva, M. Andrade, and S. Ferreira. Influence of segment length on the fitness of multivariate crash prediction models applied to a Brazilian multilane highway. *IATSS Research*, 2021.

[60] R. Yu, Mohammed Quddus, X. Wang, and K. Yang. Impact of data aggregation approaches on the relationships between operating speed and traffic safety. *Accident Analysis and Prevention*, 2018.

[61] S. Rubaiyat Aziz and S. Dissanayake. A comparative study of newly developed kansas-specific safety performance functions with hsm models for rural four-lane divided highway segments. *Journal of Transportation Safety and Security*, 2021.

[62] V. Balali, A. Ashouri Rad, and M. Golparvar-Fard. Detection, classification, and mapping of u.s. traffic signs using google street view images for roadway inventory management. *Visualization in Engineering*, 2015.

[63] S. A. Alarifi, Mohamed A. Abdel-Aty, J. Lee, and Juneyoung Park. Crash modeling for intersections and segments along corridors: A bayesian multilevel joint model with random parameters. *Analytic Methods in Accident Research*, 2017.

[64] I. C. Matarage and S. Dissanayake. Quality assessment between calibrated highway safety manual safety performance functions and calibration functions for predicting crashes on freeway facilities. *Journal of Traffic and Transportation Engineering (English Edition)*, 2020.

[65] M. Macedo, E. Kohlman Rabbani, M. Maia, M. Macedo, and B. Ferreira. GIS-based methodology for crash prediction on single-lane rural highways. *Journal of Geographic Information System*, 2021.

[66] Maria-Ioanna M. Imprialou, Mohammed Quddus, and David E. Pitfield. Predicting the safety impact of a speed limit increase using condition-based multivariate poisson lognormal regression. *Transportation Planning and Technology*, 2015.

[67] S. Mitra, M. Haque, and M.J. King. Effects of access, geometric design, and heterogeneous traffic on safety performance of divided multilane highways in India. *Journal of Transportation Safety and Security*, 2017.