Scientific Research Publishing

Journal of Transportation Technologies, 2023, 13, 282-300
https://www.scirp.org/journal/jtts
ISSN Online: 2160-0481
ISSN Print: 2160-0473

Exploring Crowdsourced Hard—Acceleration and Braking Event Data for Evaluating Safety Performance of Low-Volume Rural Highways in Iowa

Shoaib Mahmud, Christopher M. Day

Department of Civil Construction and Environmental Engineering, Iowa State University, Ames, USA
Email: smahmud@iastate.edu, cmday@iastate.edu

How to cite this paper: Mahmud, S. and Day, C.M. (2023) Exploring Crowdsourced Hard—Acceleration and Braking Event Data for Evaluating Safety Performance of Low-Volume Rural Highways in Iowa. *Journal of Transportation Technologies*, **13**, 282-300.
https://doi.org/10.4236/jtts.2023.132014

Received: March 1, 2023
Accepted: April 10, 2023
Published: April 13, 2023

Abstract

There are over four million miles of two-lane roadways across the United States, of which a substantial portion is low-volume roads (LVR). Traditionally, most traffic safety efforts and countermeasures focus on high-volume high-crash urban locations. This is because LVRs cover an extensive area, and the rarity of crashes makes it challenging to use crash data to monitor the safety performance of LVRs regularly. In addition, obtaining up-to-date roadway information, such as pavement or shoulder conditions of an extensive LVR network, can be exceptionally difficult. In recent times, crowdsourced hard-acceleration and braking event data have become commercially available, which can provide precise geolocation information and can be readily acquired from different vendors. The present paper examines the potential use of this data to identify opportunities to monitor the safety of LVRs. This research examined approximately 12 million hard-acceleration and hard-braking events over a 3-months period and 26,743 crashes, including 9373 fatal injuries over the past 5-year period. The study found a moderate correlation between hard acceleration/hard-braking events with historical crash events. This study conducted a hot spot analysis using hard-acceleration/hard-braking and crash datasets. Hotspot analysis detected spatial clusters of high-risk crash locations and detected 848 common high-risk sites. Finally, this paper proposes a combined ranking scheme that simultaneously considers historical crash events and hard-acceleration/hard-braking events. The research concludes by suggesting that agencies can potentially use the hard-acceleration and hard-braking event dataset along with the historical crash dataset to effectively supervise the safety performance of the vast network of LVRs more frequently.

Keywords

Connected Vehicle, Low Volume Highway, High-Risk Crash Sites, Hard Acceleration and Braking Events, Geographic Information System

1. Introduction

1.1. Background

Each year approximately 35,000 fatal crashes occur on roads in the United States, and over 45 percent of all crashes occur along rural roads [1]. There are over four million miles of two-lane roadways across the country, of which three million are rural. Of these, a substantial portion is low-volume roads (LVRs), where the annual average daily traffic (AADT) is less than 400 vehicles per day (VPD) [2]. Historical crash data show that such roads have higher crash rates than others. In 2019, the crash fatality rate per 100 million vehicle miles traveled (VMT) was about 1.9 times higher in rural areas compared to urban areas [1].

Traditionally, most efforts addressing traffic safety issues, particularly by the engineering and enforcement communities, have been targeted at high-volume high-crash urban locations. The primary reason is that LVRs cover an extensive area, and the conventional approach of identifying the worst-performing locations is not as effective for managing a widely distributed network. Furthermore, when allocating resources to crash mitigation strategies, it is more reasonable to emphasize higher-volume urban roads and intersections with many more crashes than LVRs. Therefore, there is a need to conduct macro-level evaluations of the safety performance of LVRs for future planning and efficient resource allocation for taking crash mitigation countermeasures.

Crash report data is widely used to detect high-crash sites in low-volume rural areas. In addition, crash report data has also been used to prioritize safety improvement measures in many locations [3] [4] [5] [6] [7]. However, the rarity of crashes on LVRs and irregularity in crash narratives make it intricate to use crash data to monitor LVR safety performance [8] [9] [10]. In addition, obtaining up-to-date roadway information, such as pavement conditions, pavement markings, or shoulder conditions of an extensive LVR network, can be exceptionally challenging. On the contrary, hard-braking (HB)/hard acceleration (HA) event data can be readily acquired from commercial providers with precise time and geolocation information. In recent studies, researchers have used such data for safety performance evaluations of intersections, work zones, and urban networks [11] [12] [13]. The present paper examines the potential use of this data to identify prospects to improve the safety and operational performance of LVRs.

1.2. Objectives

The primary objective of this study is to evaluate the relationship between the frequency of HA/HB events and the occurrence of crashes on rural roads in Iowa. In addition, this study attempts to identify crash hotspots and construct a combined ranking system using the HB and HA events for frequent evaluation and monitoring of rural roads in Iowa.

1.3. Literature Review

The most common approach for safety performance measurement of crashes and detecting high-risk crash sites is analyzing historical crash data. Crash data analyses can be categorized into two types: crash frequency analysis and crash severity analysis [14]. The general approach of crash modeling is to associate the historical crash events with other variables, such as roadway geometric properties, traffic characteristics, environmental variables, crash-associated vehicle attributes, and motorist characteristics. Three extensive survey papers comprehensively illustrate recent crash modeling techniques using historical crash data, which interested readers may look up for in-depth understanding [15] [16] [17].

However, historical crash data can only detect risks after crashes have occurred. Especially in the case of LVRs, crashes are scarce events, and crash counts may not be adequate to detect and evaluate high-risk crash sites at regular intervals. Furthermore, the lack of sufficient documentation describing the factors contributing to a crash and frequent time lags in crash reporting is quite common in historical crash data. Despite these limitations, detecting areas with excessively high crash rates or crash hot spots aids transportation agencies in identifying where safety improvements are needed [18] [19] [20].

Over the past few years, multiple commercial providers have begun to market connected vehicle (CV) data aggregated from auto manufacturers and several other sources. Although such data represent a relatively small sample of the total vehicle fleet, the data is available at much lower latency than crash report data, which may enable transportation agencies to shift from a reactive to a proactive response strategy. Vehicle geolocation data obtained from onboard mobile devices, such as smartphones and navigation aids, has been primarily used so far to develop performance measures of intersections and surface street networks. Researchers have recently begun to explore relationships between speed and braking event data and historical traffic safety data [21]-[26].

Past studies have examined braking behavior. The frequency of HB events by distracted drivers was identified to significantly influence the likelihood of rear-end collisions in a simulated driving environment [27]. In the United Kingdom, investigations using truck telematics data on harsh braking incidents instead of truck crash counts have shown significant promise in detecting potential high crash-risk sites [28]. In another study using vehicle telematics data in Atlanta, Georgia, researchers noticed that drivers involved in a crash are more likely to engage in hard-deceleration events than those not involved in crashes [29]. An investigation of the frequency of adverse driving maneuvers (e.g., sudden lane changes, extreme deceleration, etc.) showed that such events could be an effective surrogate measure for crash risk [30]. Some other studies have also explored associations between driver performance, fatigue, braking, and crash severity [31] [32]. A recent study analyzed and aggregated CV-based HB events data in 23 work zones across Indiana. The study showed a high correlation between crash events with HB events and concluded that HB events could potentially surrogate historical crash count data for prioritizing safety measures for transportation system elements [11].

Iowa developed a Comprehensive Highway Safety Plan (CHSP) in 2006. At that time, an inspection of Iowa crash data for local rural roads portrayed a detailed scenario of crashes on LVRs, ranked local roads using crash count data, and distinguished several causes of collisions on LVRs [7]. A limitation of this approach is that agencies must wait for new crash records to grow to reassess the high-risk sites and the impact of the countermeasures. In this study, we intended to identify the relationship between the commercially available HA/HB events dataset and the historical crash dataset. The intention is to utilize those event data to monitor Iowa's LVRs at regular intervals. This study analyzed three months of HA/HB events and five years of historical crash incidents conducted on LVRs in Iowa.

2. Review of Data Sources

2.1. Low-Volume Roads of Rural Iowa

LVRs are defined in Part 5 of the 2009 *Manual on Uniform Traffic Control Devices* (MUTCD) as roads outside incorporated areas with a traffic volume under 400 vehicles per day. LVRs exclude freeways or other roads that are part of state highway systems, regardless of the traffic volume [2]. The present study adopts this definition. Iowa has more than 110,000 miles of roads, many of which are in rural areas with low volumes. Previous studies have shown that crashes on LVRs constitute about half of all crashes on Iowa roads and represent a substantial portion of fatal and injury crashes [7].

The Iowa Department of Transportation (Iowa DOT) maintains records of the entire primary road network of Iowa, which can be accessed from an open web-based environment named Roadway Asset Management System (RAMS). The descriptions of the LVRs and their geometric properties are obtained from the RAMS service. Later, HA/HB events data from the commercial vendors are combined with the LVRs using the spatial join tool of ArcGIS. This study considered the LVRs where at least one hard braking or acceleration record was found. Approximately 63,000 miles of low-volume rural roadways, including 50,662 road segments, were identified for the final analysis. The created LVR network is used for further research, as shown in **Figure 1**.

2.2. Statewide Crash Data

To analyze the relationship between HA/HB events and LVR crashes, it was necessary to integrate the data into a single analysis framework, along with network element attributes. To do so, a segmentation scheme used by Iowa DOT was employed for the analysis. Five years of crash data from 2016 to 2021 were

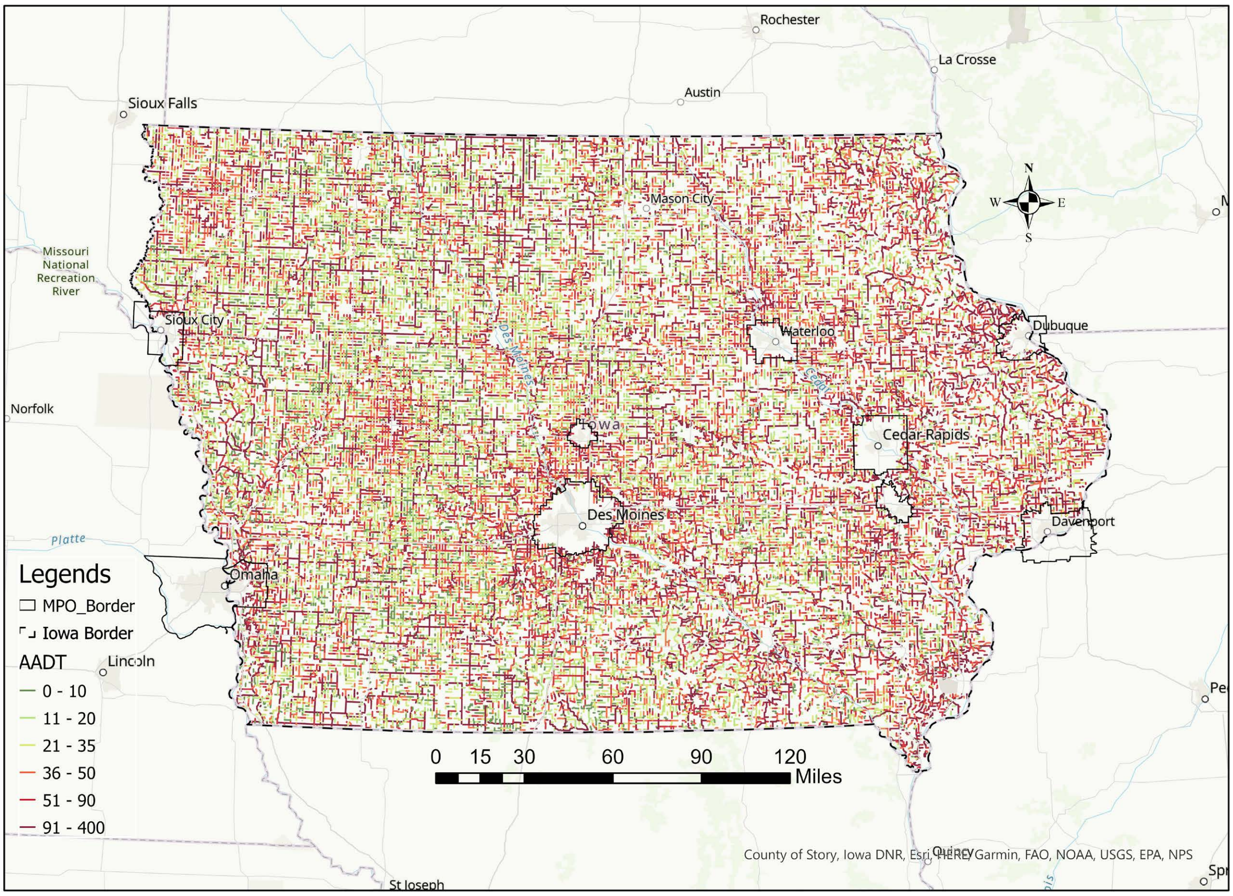

Figure 1. Low volume road network of Iowa.

obtained from the Iowa Crash Analysis Tool (ICAT). ICAT provides crash details, including driver and vehicle information, crash location geographic coordinates, crash severity, and roadway information.

At first, all five years of crashes in Iowa were mapped in ArcGIS. Then, crashes occurring within 100 ft of the LVR network were spatially joined to the nearest segment using the ArcGIS spatial join tool. Using this process, 26,743 crashes, including 9373 fatal injuries, were identified and assigned to the closest segment. One of the significant obstacles to using crash data as a performance measure for LVRs is that crashes are often underreported or reported after a considerable delay. Each individual crash was treated as one observation for each road segment rather than attempting to measure the total number of injuries. The entire five-year crash count was then divided by the segment length and the number of years for further analysis, as follows:

$$\text{Crashes/mile/year} = \frac{\text{Total no. crashes from 2016 to 2021 for each LVR segments}}{\text{Length of the LVR segment} * \text{no. of years}} \quad (1)$$

2.3. Hard Acceleration and Braking Events Data on Low-Volume Roads

Hard braking or acceleration can be identified from vehicle movement data. For this study, commercially available data was obtained wherein any acceleration or deceleration greater than 8.76 ft/s2 was used to identify a hard braking or acceleration event. The data included the geographic coordinates of the event, the timestamp, and the vehicle heading and speed. About 23 million such events were recorded in Iowa for October, November, and December 2021. **Figure 2** shows a map that illustrates the widespread geographic distribution of HA/HB events in Iowa.

The spatial join tool was used to associate hard-braking with the LVR network by filtering only those events that occurred within 100 ft of an LVR segment. More than 12 million hard-braking events were identified during the three-month study period. Comparing the number of events per mile with segment AADT reveals that hard-braking events are proportionately higher on higher-volume urban roads compared to LVRs. Hard-braking events for each segment occurring during the study period were totaled and divided by segment length and the number of months for further analysis.

$$\text{HB or HA events/mile/month} = \frac{\text{Total HB/HA events from October 2021 to December 2021}}{\text{Length of the LVR segment} * \text{no. of months}} \quad (2)$$

The summary statistics of the LVR network, along with hard-braking and crash count data, are shown in **Table 1**.

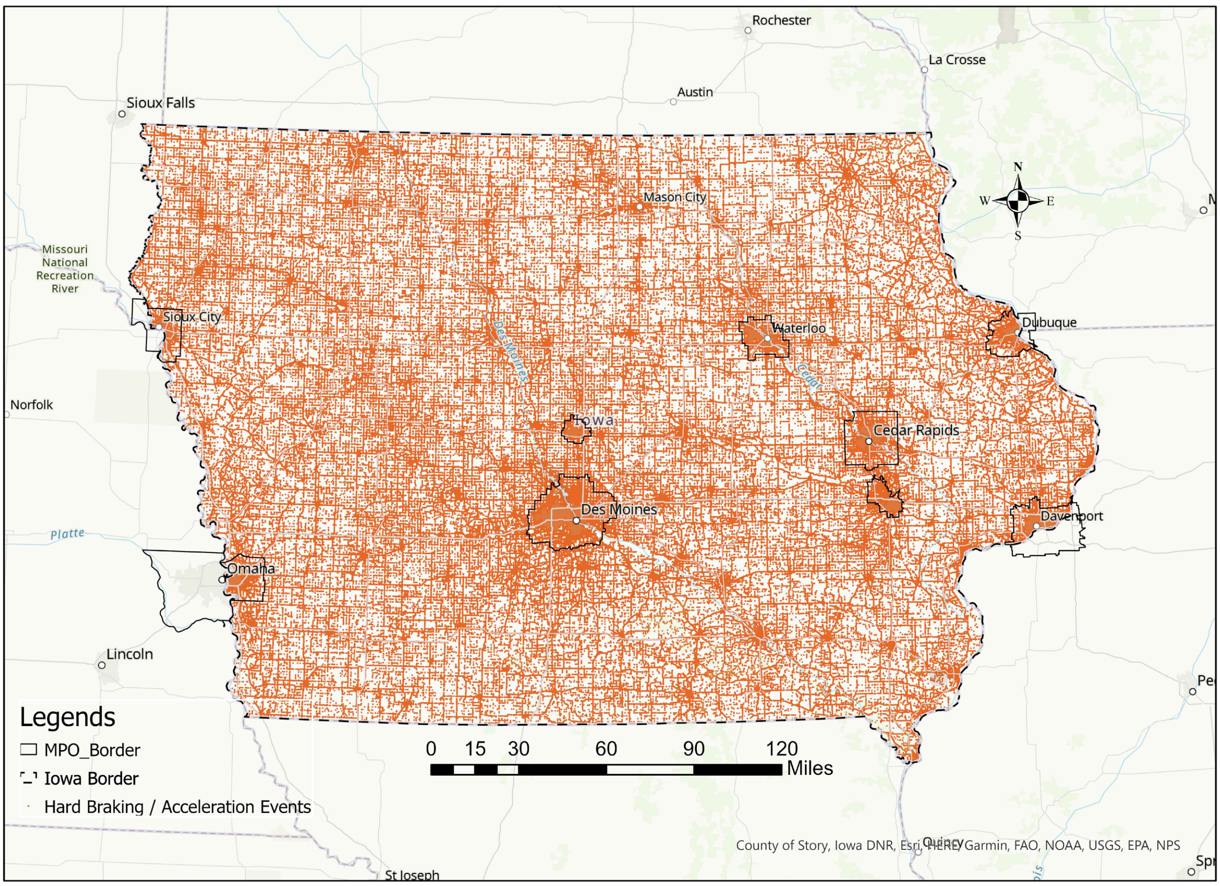

Figure 2. Geographic distribution of hard-braking events data across Iowa.

Table 1. LVR network summary statistics.

	Median	Maximum	Minimum	Standard Deviation
Segment Length (mile)	1.32	13.23	0.0124	0.873
Speed Limit (mph)	55	70	10	4.676
AADT (Vehicles per day)	45	400	5	84.641
Surface Width (ft)	24	40	8	2.68
Number of Lane	2	4	1	0.1129
Hard Braking/Acceleration Event	6	4206	2	83.645
Crash Count	0	42.0000	0	1.117
Crash Count Per Mile Per Year	0	151.98	0	2.33
Event Count Per Mile Per Year	2.24	45643.82	0.08163	469.706
Segment Length (mile)	1.32	13.23	0.0124	0.873

3. Methods

3.1. Correlation Analysis of Historical Crash Count and Hard-Breaking Events

As previously mentioned, three months of crowdsourced HA/HB event data and five years of crash data were gathered to investigate the degree to which they correlate. The following expression describes the correlation coefficient of the quantitative variables:

$$r = \left(n\sum xy - \sum x \sum y\right) \Big/ sqrt\left(\left(n\sum x^2 - \left(\sum x\right)^2\right)\left(n\sum y^2 - \left(\sum y\right)^2\right)\right) \tag{3}$$

where r is the correlation coefficient, $\sum x$ and $\sum y$ are the sums of the two associated quantitative variables, and $\sum xy$ is the summation of the product of the variables.

In multiple previous studies, linear regression was used to analyze commercially available probe datasets with different datasets of traffic events [11] [33] [34] [35]. This study also used linear regression to analyze the correlation between HA/HB events and historical crash events. The dependent variable was the aggregated number of all crashes per mile per year for each LVR road segment. Eight independent quantitative variables were obtained from the HA/HB events and roadways properties datasets. In addition to these quantitative variables, "types of terrain" was added as a categorical independent variable in the analysis. Regression models were then estimated. The following expression describes a general multiple linear regression model:

$$y = \beta_0 + \beta_2 x_1 + \beta_1 x_2 + \cdots + \beta_k x_k + \varepsilon \tag{4}$$

where y is the dependent variable, x_i are the independent variables, β_i are the coefficients, and ε is a constant.

A forward stepwise regression procedure was followed to attain the optimal equation, which includes all the significant independent variables. The forward selection process started with fitting an intercept-only model with historical crash data. Then all other independent variables were added to the initial model iteratively, and the variable with the smallest p-value was selected. New variables were added to the model until any independent variables remained significant. The stepwise regression was followed by carrying out a multicollinearity test.

The correlation between injuries and fatal crashes with the hard-braking events was also tested in the method mentioned above. For each case, the results of the regression model for three different scenarios are presented in the analysis section: the model with only HA/HB events, the model with HA/HB events and one set of categorical variables, the model with HA/HB events with the best set of independent variables.

3.2. Hot Spot Analysis of Historical Crash Count and Hard-Breaking Events

In addition to exploring correlations between driving events and crashes, the present study also seeks to identify the geographical locations of high-risk crash sites using the connected vehicle data and compare the results with the traditional method of identifying high-risk areas using historical crash data. This research applied Getis-Ord (G_i^*) spatial statistics, a common hot spot analysis technique [36] [37] [38], to determine high-risk crash sites from the historical crash count and HA/HB event database. In this study, the attributes used to define hot spots are crash count per mile per year and HA/HB event count per mile per year. The G_i^* spatial statistics can reveal clusters of high and low concentrations of spatial dependencies, which can be termed hot spots or cold spots [39].

The G_i^* spatial statistics procedure returns a z-score for each spatial feature of the dataset. Here, the method examined each LVR segment within the context of its adjacent segments and indicated the spatially significant hotspots. Therefore, any isolated segment with a high HA/HB event or crash count was not detected as a hotspot throughout the process. The analysis was performed, and the visualizations were generated using ArcGIS pro. In the visualizations, the spatial clusters of the HA/HB events and crash events were represented in confidence bins −99.9%, 99%, and 95%. The G_i^* statistics were calculated using the equations as follows [39]:

$$G_i^* = \frac{\sum_{j=1}^{n} w_i x_i - \overline{X} \sum_{j=1}^{n} w_{i,j}}{S\sqrt{\frac{\left[n\sum_{j=1}^{n} w_{i,j}{}^2 - \left(\sum_{j=1}^{n} w_{i,j}\right)^2\right]}{n-1}}} \tag{5}$$

$$\overline{X} = \frac{\sum_{j=1}^{n} x_j}{n} \tag{6}$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - \bar{X}^2} \tag{7}$$

where G_i^* is the spatial autocorrelation statistic of an event i over n events, the term x_i and x_j is the attribute value (HA/HB event rate and crash event rate) for segment i and j, $w_{i,j}$ is the spatial weight between segment i and j.

3.3. Ranking of High-Risk LVR Segments

An earlier study regarding crash risk site identification of rural roads mentioned several measures such as crash count, crash severity, crash rate, and crash quality control locations as indicators to identify high-risk crash sites. Ksaibati & Evans (2009) ranked the road segments in Wyoming using crash rate, crash per mile, fatal and injury crash mile, and equivalent property damage only (EPDO) [3]. These studies emphasized the actual number of crashes as the primary measure of high-risk site evaluation. The importance of ranking is that it benefits agencies to quickly evaluate the roadways' safety conditions and prioritize the utilization of resources to take engineering and enforcement measures. Nevertheless, the limitation of using historical data is it takes time to develop crash history on a stretch. Taking this into account, Ksaibati & Evans (2009) introduced a combined ranking where fifty percent of the weight was assigned to a road segment based on field evaluation rank. With commercially available connected vehicle data, we propose to rank the road segments of the LVR network based on HA/HB events which eventually facilitates evaluating the LVR network's safety more frequently. In the later section of this study, we present the ranking of LVR using HA/HB events and a comparison of the proposed ranking with the traditional crash count-based ranking to demonstrate an overall safety condition.

4. Comparison of Hard-Braking Events and Historical Crash Data

4.1. Correlation Analysis

For the LVRs of Iowa, the correlation among the numerical variables is illustrated using the correlation matrix shown in Figure 3. Each box in the figure characterizes a comparison between the designated quantitative variables. A high correlation coefficient indicates a strong linear relationship between the two quantitative variables. The positive and negative signs of the correlation coefficient suggest whether the associated variables are positively or negatively correlated. The correlation coefficient of 0.59 (Figure 3) indicates that the average HA/HB events per mile are moderately correlated with crash events per mile per year. The correlation matrix also indicates a fair correlation (0.68) between HB events and HA events.

Table 2 presents three different multiple linear models of crash rates with varying combinations of variables. Model 1 includes only the rate of HB/HA events.

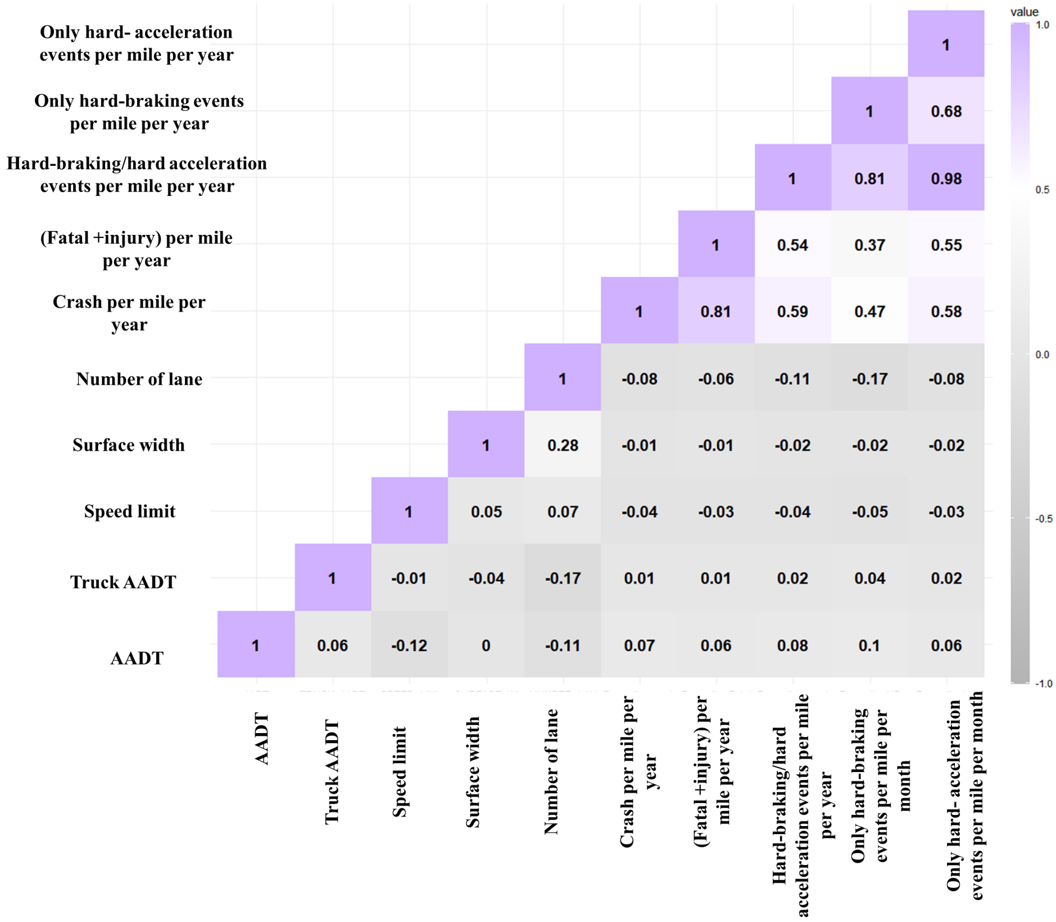

Figure 3. Correlation matrix of the quantitative variables.

Table 2. Regression model outputs for all crashes.

	Dependent Variable		
	Number of Crashes Per Mile Per Year		
Independent Variables	**(1)**	**(2)**	**(3)**
HA/HB events per mile per year	0.002925***	0.002908***	0.002905***
AADT	na	0.000695***	0.000682***
Number of lanes	na	−0.1936**	−0.1639*
Speed limit	na	−0.00574**	−0.005746**
Terrain: Flat	na	na	−0.31*
Terrain: Rolling	na	na	−0.3157*
Terrain: Hilly	na	na	−0.3924*
Constant	0.1577***	0.8001***	1.077***
Observations	50,662	50,662	50,662
Adjusted R^2	0.3464	0.3474	0.3476
Residual Standard Error	7.352	1.871	1.885
F-Statistic	26,860***	6743***	3856***

*p < 0.1; **p < 0.05; ***p < 0.01; na = not applicable.

The coefficient is estimated as 0.002925, corresponding to one crash per mile per year occurring for every 342 HA/HB events per mile per year. Models 2 and 3 introduce additional variables associated with volume and geometry. Although they are statistically significant, the inclusion of these variables did not result in much change to the HA/HB rate coefficient, nor did they change the R^2 value. However, the value of the constant term is changed with the addition of the variables. This suggests that the relationship between the rate of HA/HB events and crashes is independent of the other variables.

A similar analysis was carried out for fatal/injury crashes. The data is presented in **Table 3**. Here, the coefficient of Model 1 corresponds to one fatal/injury crash per 983 HA/HB events per mile per year. All the models developed here had p-values of less than 0.01. However, these models show minor improvements due to the inclusion of other associated quantitative and categorical variables. This implies that although some statistically significant variables exist, such as the number of lanes or speed limit, their effect size on predicting the number of crashes is not meaningful.

4.2. Hot Spot Analysis

The correlation analysis indicates that the HA/HB events per mile are moderately correlated with crashes per mile. However, it is also critical to locate the geographical location of high-risk clusters using hard-braking events and compare the results with identifying high-risk areas using historical crash data. As mentioned earlier, G_i^* spatial statistics were used to identify high-risk locations. The results of the hotspot analysis are presented in **Figure 4**, which reveals

Table 3. Regression model outputs for fatal & injury involved crashes.

	Dependent Variable		
	Number of Fatal/Injury Crashes Per Mile Per Year		
Independent Variables	**(1)**	**(2)**	**(3)**
HA/HB events per mile per year	0.001017***	0.001014***	0.001014***
AADT	na	0.000236***	0.000234***
Number of lanes	na	−0.01754	−0.1601
Speed limit	na	−0.002049**	−0.002079**
Terrain: Flat	na	na	−0.0972
Terrain: Rolling	na	na	−0.08490
Terrain: Hilly	na	na	−0.0725
Constant	0.04877***	0.1068***	0.02368
Observations	50,662	50,662	50,662
Adjusted R^2	0.2874	0.2881	0.2882
Residual Standard Error	0.7523	1.871	1.885
F-Statistic	20,440***	5125***	2930***

*p < 0.1; **p < 0.05; ***p < 0.01; na = not applicable.

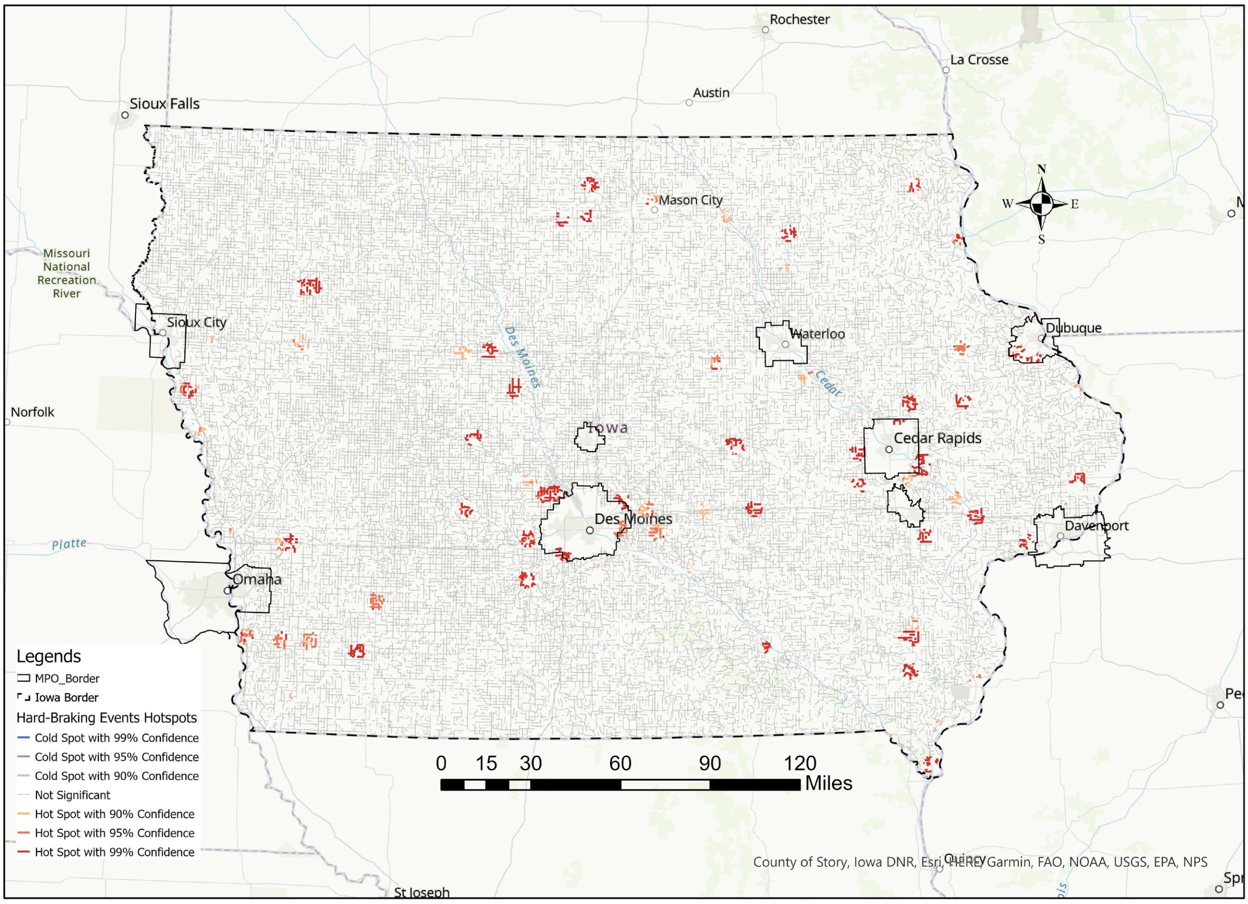

Figure 4. Hotspots of high-risk sites detected using hard-braking events.

hotspots of HA/HB events are dispersed all over Iowa. In total, 1659 (3.27%) segments are identified as hotspots from over fifty thousand LVR segments of the entire network. There appears to be a concentration of the hotspot location near Central Iowa, the Southwest part of Iowa, and the Cedar Rapids Area.

The fundamental goal of this study is to evaluate if HA/HB events can surrogate and complement historical crash count data to detect high-risk sites. Therefore, it is expected that hotspot analysis would display a substantial number of mutual high-risk segments for both datasets. Figure 5 shows the hotspots identified using historical crash data. Figure 6 shows the hotspots identified using both HA/HB events and historical crash count datasets. In total, 2,371 (4.68%) segments are identified as high-risk crash hotspots using the crash count dataset (Figure 5). Among them, 848 segments (Figure 6) show both the hard-braking event and historical crash count datasets, which is nearly 50% of the high-risk segments identified by hard-braking events. This leads to the inference that hard-braking data can potentially identify high-risk crash locations and help prioritize safety investments in LVRs for agencies.

4.3. Ranking of LVRs

The third objective of this study is to rank LVR segments by safety performance.

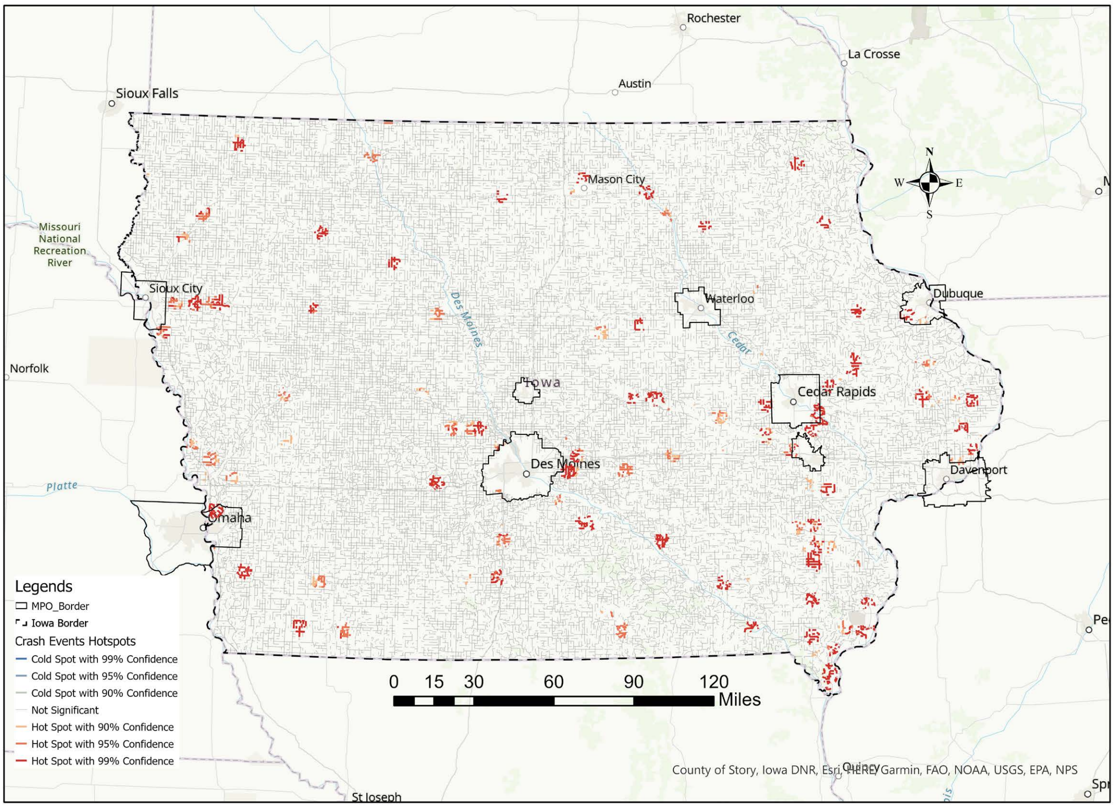

Figure 5. Hotspots of high-risk sites detected using historical crash data.

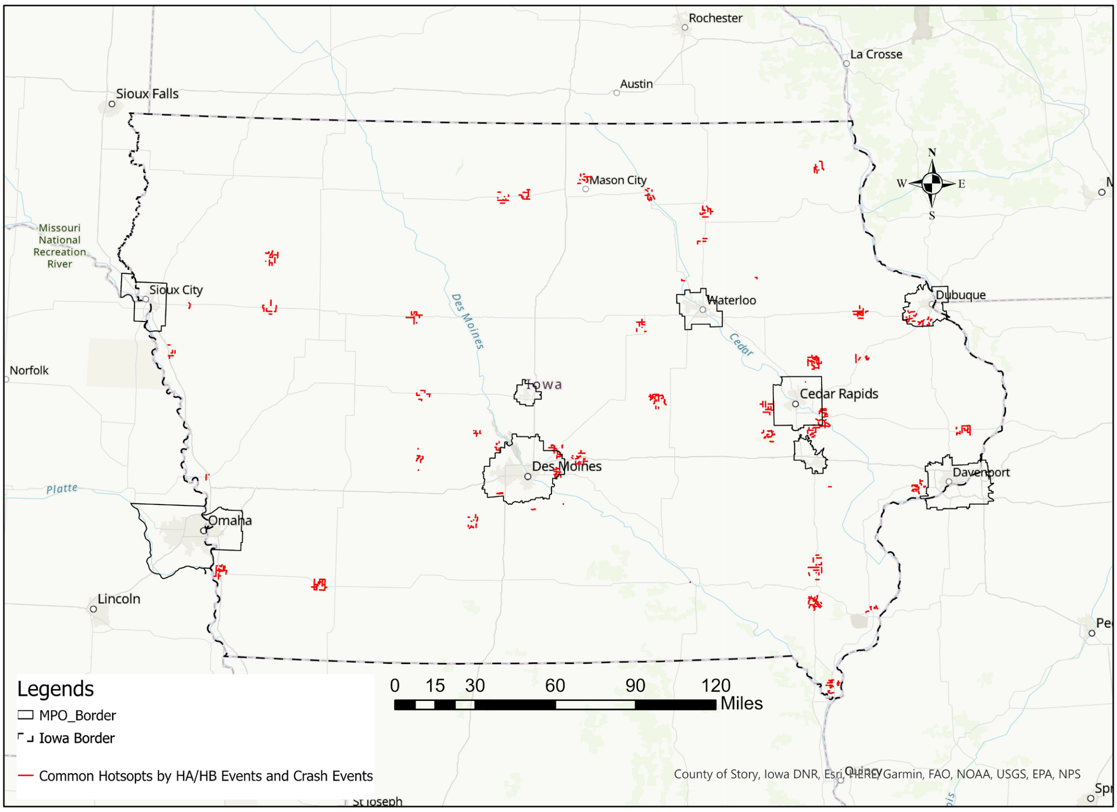

Figure 6. Common hotspots of high-risk sites.

To do so, both the HB/HA events and crash counts were used. Historical crash performance is typically used to report safety performance [3] [7]. However, some researchers have investigated the safety performance of LVRs by combining crash data with other surrogate safety measures. For example, Ksaibati & Evans (2009) introduced a combined ranking system for rural roads in Wyoming, where fifty percent of the weight was allocated to a road segment based on a field evaluation score. We propose a novel combined ranking scheme in the present study, where fifty percent of the weight is assigned to the HA/HB events based ranking score [3]. The combined ranking score is defined by the following equation:

$$\text{Combined Ranking} = \text{Crash Ranking} * 50\% + \text{HA/HB event based Ranking} * 50\% \quad (8)$$

Table 4 shows the list of the top 20 high-risk LVR segments. The individual values and ranks using crash and hard-braking event datasets are also presented for each segment (Table 4). It can be noted that the respective ranks using both datasets are not far from each other, which is expected. This ranking makes it possible to prioritize high-risk locations according to their previous crash history and recent safety performance, with those at the top of the list needing further

Table 4. The Top 20 High-Risk LVR segments according to combined rank.

Combined Rank	Road Name	County	Rank by Crash	Crash/mi/year	Rank by HB Event	Events/mile/month
1	Cox Springs Road	Dubuque	2	124.00	1	45643.82
2	Highway T47	Tama	3	111.85	5	19813.69
3	Northeast 94th Ave.	Polk	7	101.63	2	40651.33
5	F Avenue	Tama	5	104.21	7	17427.05
5	Zachary Avenue	Muscatine	9	89.57	3	38771.70
5	Iowa Avenue	Henry	4	111.12	8	15278.83
7	Pitlik Drive	Linn	14	70.48	13	10267.08
8	Argyle Raod	Lee	11	75.93	18	8528.61
9	Us Highway 69	Hancock	19	56.76	14	10097.86
10	330th Avenue	Clinton	13	74.26	21	8284.83
11	355th Street	Lee	16	61.69	22	8276.26
12	32 Avenue	Benton	29	46.79	11	12018.25
13	O Avenue	Montgomery	34	44.10	16	8599.13
14	300th Street	Cerro Gordo	6	103.98	46	3572.75
15	Bradley Court	Linn	38	41.41	15	9885.48
16	Jecklin Lane	Dubuque	37	41.73	20	8345.67
17	130th Street	Des Moines	22	50.51	37	4391.57
18	275th Street	Bremer	25	47.51	40	3933.11
20	Highway 977	Cherokee	40	39.79	27	6750.27
20	Us 34	Mills	33	44.56	34	4961.26

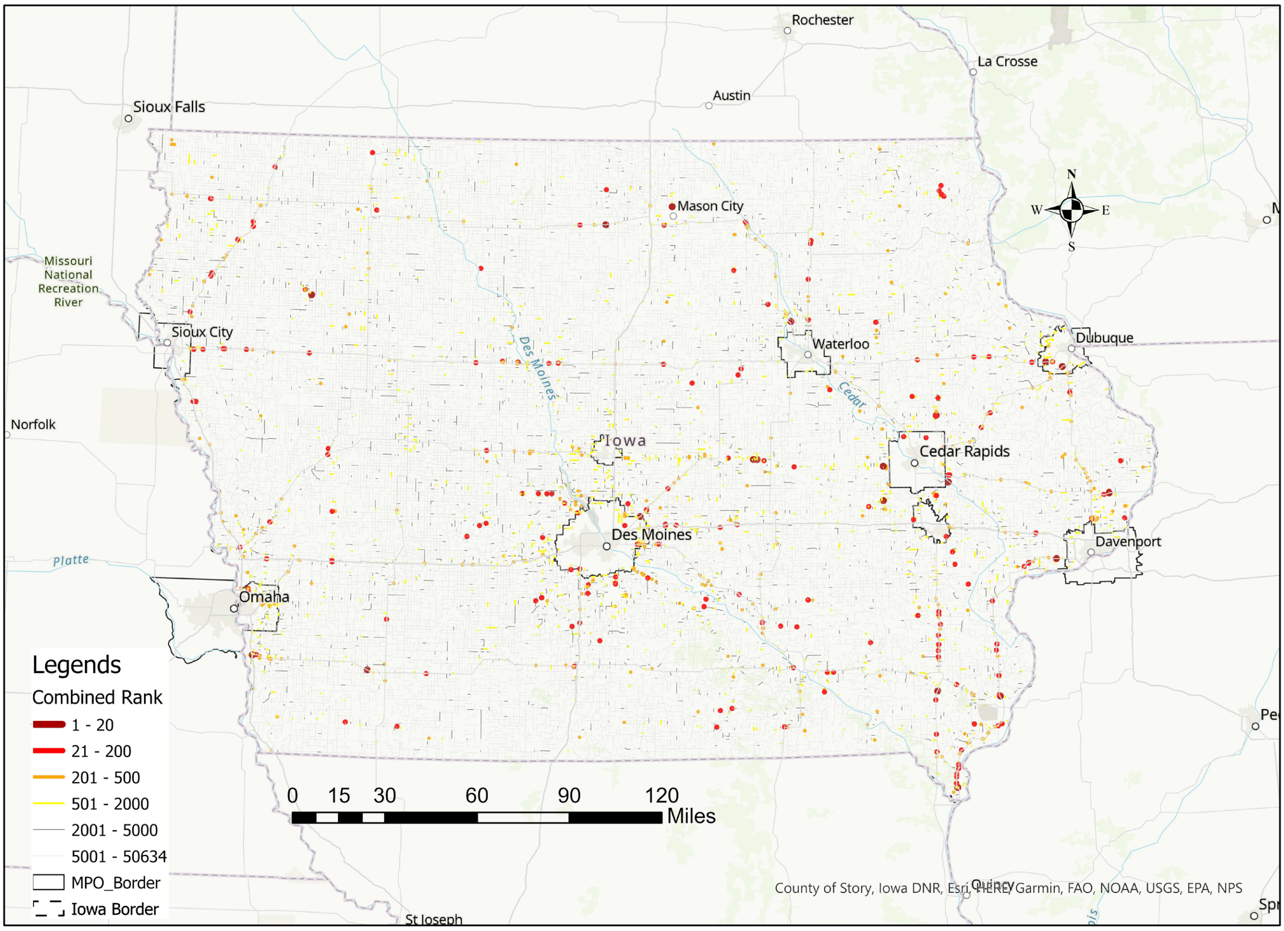

Figure 7. High-risk sites by combined rank.

evaluation to take effective countermeasures. Essentially, combined ranking addresses the issue that historical crash data requires time to develop and consider the recent performance of the roadways.

The combined rank of all the road segments is presented using a color-coded map in Figure 7. Though the high-risk location is distributed all over Iowa, many sites are detected near Central Iowa and Cedar Rapids. Because of the randomness of LVR crashes, keeping such an extensive network under constant monitoring is extremely challenging. The developed list of high-risk crash locations using the hard-braking events provides an effective monitoring tool and the candidates for further safety-related data collection.

5. Conclusions

The results of this study demonstrate that data from commercial connected vehicle data are useful for analyzing the safety performance and visualizing the high-risk crash sites of LVRs. The study compared the rates of HA/HB events with crash rates using three months of data for a 63,000-mile LVR network in Iowa. Three separate analyses were undertaken as part of the study: a correlation analysis, hotspot analysis, and ranking of LVR segments.

The correlation analysis indicates a moderate relationship between the rate of HA/HB events and crashes. The hotspot analysis effectively visualizes high-risk areas using hard-braking and crash datasets. Hotspot analysis portrays 848 LVR segments are identified as the common high-risk locations by both hard-braking events and crash datasets. This indicates that hard-braking events can effectively surrogate crash count data for monitoring potential crash risk. The study proposes a new combined ranking that considers historical crash count and HA/HB events simultaneously. The purpose of ranking is to assist agencies in conducting informed and economical field evaluations to detect causative factors and take necessary engineering and enforcement countermeasures.

This preliminary analysis offers insight into how CV data could be integrated into site selection and prioritization methods, focusing on the safety of LVRs. This study detected some high-risk sites in Iowa (**Table 4**) where prompt enforcement actions might be taken to address the immediate safety risk in these locations. In addition, after further assessment, engineering efforts should be taken in these locations for long-term safety improvements. The usage of this data-driven method for systematic and frequent evaluation of low-volume roads is a new approach to rural road safety. This approach can provide awareness of the safety of low-volume roads and aid in identifying areas that require rigorous monitoring or future improvement. Future research could explore the impact of the amount of HA/HB event data to determine the amount needed and whether seasonal effects need to be accounted for when sampling the CV data. Further, the ranking methodology presented here combines crash data and CV data with equal weighting, and additional study is needed to assess this approach. Finally, future work will seek to expand the analysis to incorporate other types and classifications of roadways.

Author Contribution Statement

C.M. Day provided guidance on analysis and helped draft the paper. S.M. carried conceived the study, out the analysis, prepared the initial draft of the paper, and created the figures.

Acknowledgements

This work was supported in part by the Iowa Department of Transportation. CV data used in this study were obtained from Wejo, Ltd. The contents of this manuscript represent the views of the authors, who are responsible for the facts and the correctness of the data presented herein, and do not necessarily reflect the official views or policies of any organizations.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] National Highway Traffic Safety Administration (2021) Traffic Safety Facts 2019

Data. National Center for Statistics and Analysis, Washington, DC. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813336

[2] MUTCD (2009) Manual on Uniform Traffic Control Devices. Part 5—Traffic Control Devices for Low-Volume Roads. FHWA, U.S. Department of Transportation, Washington DC. https://mutcd.fhwa.dot.gov/pdfs/2009r1r2/part5.pdf

[3] Ksaibati, K. and Evans, B. (2009) Wyoming Rural Road Safety Program. 88*th Annual Meeting of the Transportation Research Board*, Washington DC, 11-15 January 2009, 49-51. https://onlinepubs.trb.org/onlinepubs/trnews/trnews273rpo.pdf

[4] Alrejjal, A., Moomen, M. and Ksaibati, K. (2022) Evaluating the Effectiveness of Law Enforcement in Reducing Truck Crashes for a Rural Mountainous Freeway in Wyoming. *Transportation Letters*, **14**, 807-817. https://doi.org/10.1080/19427867.2021.1949089

[5] Sharafeldin, M., Farid, A. and Ksaibati, K. (2022) Investigating the Impact of Roadway Characteristics on Intersection Crash Severity. *Eng*, **3**, 412-423. https://doi.org/10.3390/eng3040030

[6] Mahmud, S. and Day, C.M. (2022) Leveraging Data-Driven Traffic Management in Smart Cities: Datasets for Highway Traffic Monitoring. In: Alavi, A.H., *et al.*, Eds., *The Rise of Smart Cities*, Butterworth-Heinemann, Oxford, 583-607. https://doi.org/10.1016/B978-0-12-817784-6.00009-6

[7] Souleyrette, R.R., Caputeu, M., McDonald, T.J., Sperry, R.B., Hans, Z.N. and Cook, D. (2010) Safety Analysis of Low-Volume Rural Roads in Iowa. Institute for Transportation, Iowa State University, Ames. https://publications.iowa.gov/21127/1/IADOT_InTrans_07_309_Safety_Analysis_Low_Volume_Rural_Roads_Iowa_2010.pdf

[8] Janstrup, K.H., Kaplan, S., Hels, T., Lauritsen, J. and Prato, C.G. (2016) Understanding Traffic Crash Under-Reporting: Linking Police and Medical Records to Individual and Crash Characteristics. *Traffic Injury Prevention*, **17**, 580-584. https://doi.org/10.1080/15389588.2015.1128533

[9] Yamamoto, T., Hashiji, J. and Shankar, V.N. (2008) Underreporting in Traffic Accident Data, Bias in Parameters and the Structure of Injury Severity Models. *Accident Analysis & Prevention*, **40**, 1320-1329. https://doi.org/10.1016/j.aap.2007.10.016

[10] Watson, A., Watson, B. and Vallmuur, K. (2015) Estimating Under-Reporting of Road Crash Injuries to Police Using Multiple Linked Data Collections. *Accident Analysis & Prevention*, **83**, 18-25. https://doi.org/10.1016/j.aap.2015.06.011

[11] Desai, J., Li, H., Mathew, J.K., Cheng, Y.-T., Habib, A. and Bullock, D.M. (2021) Correlating Hard-Braking Activity with Crash Occurrences on Interstate Construction Projects in Indiana. *Journal of Big Data Analytics in Transportation*, **3**, 27-41. https://doi.org/10.1007/s42421-020-00024-x

[12] Zhang, S. and Abdel-Aty, M. (2022) Real-Time Crash Potential Prediction on Freeways Using Connected Vehicle Data. *Analytic Methods in Accident Research*, **36**, Article ID: 100239. https://doi.org/10.1016/j.amar.2022.100239

[13] Hunter, M., Saldivar-Carranza, E., Desai, J., Mathew, J.K., Li, H. and Bullock, D.M. (2021) A Proactive Approach to Evaluating Intersection Safety Using Hard-Braking Data. *Journal of Big Data Analytics in Transportation*, **3**, 81-94. https://doi.org/10.1007/s42421-021-00039-y

[14] Das, S., Tsapakis, I. and Khodadadi, A. (2021) Safety Performance Functions for Low-Volume Rural Minor Collector Two-Lane Roadways. *IATSS Research*, **45**, 347-356. https://doi.org/10.1016/j.iatssr.2021.02.004

[15] Lord, D. and Mannering, F. (2010) The Statistical Analysis of Crash-Frequency Data: A Review and Assessment of Methodological Alternatives. *Transportation Research Part A*: *Policy and Practice*, **44**, 291-305. https://doi.org/10.1016/j.tra.2010.02.001

[16] Savolainen, P.T., Mannering, F.L., Lord, D. and Quddus, M.A. (2011) The Statistical Analysis of Highway Crash-Injury Severities: A Review and Assessment of Methodological Alternatives. *Accident Analysis & Prevention*, **43**, 1666-1676. https://doi.org/10.1016/j.aap.2011.03.025

[17] Mannering, F.L. and Bhat, C.R. (2014) Analytic Methods in Accident Research: Methodological Frontier and Future Directions. *Analytic Methods in Accident Research*, **1**, 1-22. https://doi.org/10.1016/j.amar.2013.09.001

[18] Wada, K., Wallner, G. and Vos, S. (2022) Studying the Utilization of a Map-Based Visualization with Vitality Datasets by Domain Experts. *Geographies*, **2**, 379-396. https://doi.org/10.3390/geographies2030024

[19] Kronenberg, C., Woodward, L., DuBose, B. and Weissman, D. (2015) Achieving Vision Zero: Data-Driven Investment Strategy to Eliminate Pedestrian Fatalities on a Citywide Level. *Transportation Research Record*, **2519**, 146-156. https://doi.org/10.3141/2519-16

[20] Montella, A. (2010) A Comparative Analysis of Hotspot Identification Methods. *Accident Analysis & Prevention*, **42**, 571-581. https://doi.org/10.1016/j.aap.2009.09.025

[21] Quddus, M. (2013) Exploring the Relationship between Average Speed, Speed Variation, and Accident Rates Using Spatial Statistical Models and GIS. *Journal of Transportation Safety & Security*, **5**, 27-45. https://doi.org/10.1080/19439962.2012.705232

[22] Banihashemi, M., Dimaiuta, M., Zineddin, A., Spear, B., Smadi, O. and Hans, Z. (2019) Using Linked SHRP2 RID and NPMRDS Data to Study Speed-Safety Relationships on Urban Interstates and Major Arterials. *The* 98*th TRB Annual Meeting*, Washington DC, 13-17 January 2019, 65-81.

[23] Birma, G.J., Agaja, S.A. and Ndu, J.C. (2022) Evaluation of Safety Culture in Institutional Chemical Analytical Laboratories in Oghara and Warri, Delta State, Nigeria. *Open Journal of Safety Science and Technology*, **12**, 17-29. https://doi.org/10.4236/ojsst.2022.121002

[24] Moreno, A.T. and García, A. (2013) Use of Speed Profile as Surrogate Measure: Effect of Traffic Calming Devices on Crosstown Road Safety Performance. *Accident Analysis & Prevention*, **61**, 23-32. https://doi.org/10.1016/j.aap.2012.10.013

[25] Mushori, J., Rambo, C.M., Wafula, C.M. and Matu, J. (2021) Evaluating Contractors' Safety Record and Its Influence on Performance of Road Infrastructural Projects. *World Journal of Engineering and Technology*, **9**, 203-228. https://doi.org/10.4236/wjet.2021.92015

[26] Ederer, D.J., Rodgers, M.O., Hunter, M.P. and Watkins, K.E. (2020) Case Study Using Probe Vehicle Speeds to Assess Roadway Safety in Georgia. *Transportation Research Record*, **2674**, 554-562. https://doi.org/10.1177/0361198120948858

[27] Haque, M.M. and Washington, S. (2015) The Impact of Mobile Phone Distraction on the Braking Behaviour of Young Drivers: A Hazard-Based Duration Model. *Transportation Research Part C*: *Emerging Technologies*, **50**, 13-27. https://doi.org/10.1016/j.trc.2014.07.011

[28] Kamla, J., Parry, T. and Dawson, A. (2019) Analysing Truck Harsh Braking Incidents to Study Roundabout Accident Risk. *Accident Analysis & Prevention*, **122**,

365-377. https://doi.org/10.1016/j.aap.2018.04.031

[29] Jun, J., Ogle, J. and Guensler, R. (2007) Relationships between Crash Involvement and Temporal-Spatial Driving Behavior Activity Patterns: Use of Data for Vehicles with Global Positioning Systems. *Transportation Research Record*, **2019**, 246-255. https://doi.org/10.3141/2019-29

[30] Musicant, O., Bar-Gera, H. and Schechtman, E. (2010) Electronic Records of Undesirable Driving Events. *Transportation Research Part F: Traffic Psychology and Behaviour*, **13**, 71-79. https://doi.org/10.1016/j.trf.2009.11.001

[31] Kiefer, R.J., LeBlanc, D.J. and Flannagan, C.A. (2005) Developing an Inverse Time-to-Collision Crash Alert Timing Approach Based on Drivers' Last-Second Braking and Steering Judgments. *Accident Analysis & Prevention*, **37**, 295-303. https://doi.org/10.1016/j.aap.2004.09.003

[32] Mollicone, D., Kan, K., Mott, C., Bartels, R., Bruneau, S., van Wollen, M., Sparrow, A.R. and Van Dongen, H.P.A. (2019) Predicting Performance and Safety Based on Driver Fatigue. *Accident Analysis & Prevention*, **126**, 142-145. https://doi.org/10.1016/j.aap.2018.03.004

[33] Emtenan, A.M.T., Haghighat, A., Shields, M., Shaw, J., Hawley, P., Sharma, A. and Day, C.M. (2022) Exploratory Regression Models for Estimating Right-Turn-on-Red Volume on Exclusive Right-Turn Lanes at Signalized Intersections. *Transportation Research Record*. https://doi.org/10.1177/03611981221116370

[34] Tahsin Emtenan, A.M. and Day, C.M. (2022) Correlation Analysis for Exploring the Relationship between Probe Vehicle Data and Event-Based Traffic Signal Performance Measures. *Transportation Research Record: Journal of the Transportation Research Board*, **2676**, 587-600. https://doi.org/10.1177/03611981221084684

[35] Day, C.M., Emtenan, T., Mahmud, S. and Knickerbocker, S. (2021) Data Driven Identification of Candidates for Operational Improvement (No. InTrans Project 19-716). https://intrans.iastate.edu/app/uploads/2021/11/data_driven_ID_of_operational_improvement_candidates_w_cvr.pdf

[36] Manap, N., Borhan, M.N., Yazid, M.R.M., Hambali, M.K.A. and Rohan, A. (2019) Determining Spatial Patterns of Road Accidents at Expressway by Applying Getis-Ord G_i^* Spatial Statistic. *International Journal of Recent Technology and Engineering*, **8**, 345-350. https://doi.org/10.35940/ijrte.C1004.1183S319 https://www.ijrte.org/wp-content/uploads/papers/v8i3s3/C10041183S319.pdf

[37] Manap, N., Borhan, M.N., Yazid, M.R.M., Hambali, M.K.A. and Rohan, A. (2021) Identification of Hotspot Segments with a Risk of Heavy-Vehicle Accidents Based on Spatial Analysis at Controlled-Access Highway. *Sustainability*, **13**, 1487. https://doi.org/10.3390/su13031487

[38] Songchitruksa, P. and Zeng, X. (2010) Getis-Ord Spatial Statistics to Identify Hot Spots by Using Incident Management Data. *Transportation Research Record*, **2165**, 42-51. https://doi.org/10.3141/2165-05

[39] Getis, A. and Ord, J.K. (2010) The Analysis of Spatial Association by Use of Distance Statistics. In: Anselin, L. and Rey, S.J., Eds., *Perspectives on Spatial Data Analysis*, Springer, Berlin, 127-145. https://doi.org/10.1007/978-3-642-01976-0_10 https://link.springer.com/chapter/10.1007/978-3-642-01976-0_10