

Article

Estimating Urban Traffic Safety and Analyzing Spatial Patterns through the Integration of City-Wide Near-Miss Data: A New York City Case Study

Chuan Xu ^{1,*}, Jingjin Gao ², Fan Zuo ² and Kaan Ozbay ²

¹ School of Transportation and Logistics, Southwest Jiaotong University, No. 111, the 1st North Section of the 2nd Ring Rd., Chengdu 610031, China

² Department of Civil and Urban Engineering, Tandon School of Engineering, New York University, 6 MetroTech Center, 4th Floor, Brooklyn, NY 11201, USA; jingjin.gao@nyu.edu (J.G.); fz380@nyu.edu (F.Z.); kaan.ozbay@nyu.edu (K.O.)

* Correspondence: xuchuan@swjtu.edu.cn

Abstract: City-wide near-miss data can be beneficial for traffic safety estimation. In this study, we evaluate urban traffic safety and examine spatial patterns by incorporating city-wide near-miss data (59,277 near-misses). Our methodology employs a grid-based method, the Empirical Bayes (EB) approach, and spatial analysis tools including global Moran's I and local Moran's I. The study findings reveal that near-misses have the strongest correlation with observed crash frequency among all the variables studied. Interestingly, the ratio of near-misses to crashes is roughly estimated to be 1957:1, providing a potentially useful benchmark for urban areas. For other variables, an increased number of intersections and bus stops, along with a greater road length, contribute to a higher crash frequency. Conversely, residential and open-space land use rates show a negative correlation with crash frequency. Through spatial analysis, potential risk hotspots including roads linking bridges and tunnels, and avenues bustling with pedestrian activity, are highlighted. The study also identified negative local spatial correlations in crash frequencies, suggesting significant safety risk variations within relatively short distances. By mapping the differences between observed and predicted crash frequencies, we identified specific grid areas with unexpectedly high or low crash frequencies. These findings highlight the crucial role of near-miss data in urban traffic safety policy and planning, particularly relevant with the imminent rise of autonomous and connected vehicles. By integrating near-miss data into safety estimations, we can develop a more comprehensive understanding of traffic safety and, thus, more effectively address urban traffic risks.

Keywords: near-misses; empirical Bayes; vulnerable road users; urban area; safety estimation



Citation: Xu, C.; Gao, J.; Zuo, F.; Ozbay, K. Estimating Urban Traffic Safety and Analyzing Spatial Patterns through the Integration of City-Wide Near-Miss Data: A New York City Case Study. *Appl. Sci.* **2024**, *14*, 6378. <https://doi.org/10.3390/app14146378>

Academic Editor: Luís Picado Santos

Received: 20 June 2024

Revised: 17 July 2024

Accepted: 19 July 2024

Published: 22 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

According to the World Health Organization, approximately 1.35 million people die in road traffic crashes annually [1], making it the eighth leading cause of death worldwide. Furthermore, an additional 20 to 50 million individuals suffer non-fatal injuries [2], often resulting in long-term disabilities. In densely populated urban areas like New York City, traffic safety is worsened by the sheer volume of vehicles, pedestrians, and cyclists that share the streets. With over 259 fatalities and 50,733 injuries reported in 2022 [3], the city faces numerous challenges, including the high cost of accidents, the vulnerability of certain road users, and the impact of crashes on traffic congestion. Addressing these issues is of utmost importance for the well-being of the city's residents, commuters, and visitors.

Fortunately, a unified commitment to promoting safety for all is demonstrated at both federal and city levels. The Federal Highway Administration's (FHWA) Strategic Plan for 2022–2026 [4] places traffic safety at the forefront of its goals, emphasizing five key components: Safety Design, Safety System, Safe Public, Safe Workers, and Critical

Infrastructure Cybersecurity. This comprehensive approach highlights the need for innovative infrastructure solutions, data-driven methodologies, public awareness, worker safety, and cybersecurity measures to address the multifaceted challenges of traffic safety. New York City's Vision Zero Plan [5], launched in 2014, is another key initiative that shares the objective of improving traffic safety. The plan aims to eliminate traffic-related fatalities and serious injuries by implementing a combination of enforcement, education, and engineering measures. Despite a nationwide rise in traffic fatalities, New York City defied the trend in 2022 with a 6.6% reduction in overall traffic fatalities, and a 6.3% reduction in pedestrian fatalities, with Vision Zero in effect. These federal and city-level initiatives spotlight the collective responsibility to prioritize traffic safety.

Historical crash data play a critical role in traffic safety analysis, as they provide a foundation for understanding the factors contributing to crashes, identifying high-risk areas, and countermeasure assessment. However, despite their value, the rarity of crashes, the issue of underreporting, and low location accuracy limit their usage in practice. Using solely crash data to estimate safety often suffers from the regression-to-the-mean problem [6,7]. The Empirical Bayes (EB) method [6,7] is a potential solution because it can combine the observed crash frequency from the real world and the expected crash frequency predicted by a crash prediction model. It is well suited for crash frequency estimation, given that it accounts for site-specific effects and regression-to-the-mean bias.

A reliable crash prediction model can be a solid foundation for the EB method. Furthermore, integrating valuable information such as near-misses [8–11] to build a crash prediction model can be a potential improvement. With the advancement of computer vision technology, devices, either via fixed infrastructures (e.g., roadside cameras) or onboard devices (e.g., in-vehicle cameras, radars), have emerged as valuable tools in detecting near-misses. These devices can capture near-miss incidents, providing a more comprehensive understanding of potential hazards and close interactions among road users that might not result in reported crashes. With the anticipated rise of autonomous and connected vehicles, the collection of near-miss data is expected to become both more feasible and cost-effective. As such, the effective analysis of these easily acquired near-miss data can yield insights that go beyond individual crashes, playing a crucial role in enhancing road safety estimations.

This study integrates city-wide near-misses to estimate urban traffic safety and analyzes its spatial patterns. To effectively organize the data, we employed a grid-based analysis that enables the integration of various factors affecting traffic safety. We first examined if a correlation exists between the crash records and near-miss data collected via in-vehicle cameras through computer vision technologies. The near-miss data were provided by the industry partner Mobileye. Then, we modeled the crash frequency by considering several variables, including near-misses, traffic volume, the number of intersections, road length, land use percentage, and population density. Then, we calculated the EB estimated crash frequency to represent traffic safety in each grid. Finally, we analyze the spatial distribution patterns, spatial autocorrelations, observation, and prediction differences.

2. Literature Review

Near-misses, also known as traffic conflicts, near-crashes, and safety-critical events [10], are growing in importance for traffic safety studies. This trend is likely driven by their shorter observation periods [12] and increased obtainability [10] due to advancements in technology.

While various definitions of near-misses exist—such as a traffic event that requires a rapid evasive maneuver [13], or situations where two or more road users come so close to each other that there is a risk of collision [14]—most studies employ the Time to Collision (TTC) metric with a specific threshold to identify these incidents. In previous studies, for example, TTC thresholds of 0.5 s [15], 1.5 s [16], 2.5 s [17], and 6 s [18] have been employed for near-miss identification. Using other time-based measures such as PET [19], Time to Collision with Disturbance (TTCD) [20], and two-dimension TTC (2D-TTC) [21] has also

been proposed. These time-based measures, employed for identifying near-misses, have demonstrated correlations with crashes [10]. However, as of now, there is no standardized method for near-miss identification.

A systematic review [10] suggests that real-world near-miss data collection methods can be categorized into road user-level and facility-level methods. The road user-level method utilizes onboard devices (OBDs) to observe near-misses [22]. The primary function of these OBDs is to measure the distance and relative speed between the ego vehicle and the forward vehicle, enabling the calculation of time-based safety measures. Sensors frequently used include cameras employing computer vision technology [23], radar [24], lidar [25], and their combinations [26]. Since these devices are vehicle-mounted, they can capture near-misses involving the ego vehicle. The sampling of near-miss events is based on the Market Penetration Rate (MPR) of OBD-equipped vehicles and their movements.

In contrast, the facility-level method involves sensors installed near specific fixed facilities, like roadside cameras [27], microwave radars [28], roadside lidar [29], and weigh-in-motion detectors [30]. This method can capture all near-misses within the detection range of the sensors. However, due to often limited sensor detection range, the facility-level method is typically unsuitable and uneconomical for collecting near-misses across wide spaces, such as city-wide road networks.

Most recent safety estimation studies leveraging near-miss data primarily focus on smaller-scale road facilities like intersections. Roadside cameras or drones are typically utilized to capture road user trajectories at intersections, from which near-misses are extracted to build extreme value models for crash risk estimation [9,10,31,32]. However, the application of this method on a larger scale, such as city-wide road networks, presents challenges. Despite near-miss data providing valuable insights for safety estimation [11], their integration into safety evaluations is seldom seen. For example, the safety performance functions provided by the Highway Safety Manual (HSM) [33] do not incorporate near-miss data. Nevertheless, the integration of real-world near-miss data into the safety evaluation of large-scale road networks presents an interesting avenue for exploration.

3. Methodology

3.1. Grid-Based Method

The grid cell-based method divides a geographical area into a grid of uniformly sized and shaped cells. Each cell represents an independent spatial unit used to collect, analyze, and display data. It has been used in various past studies [34–36]. This cell size was chosen because it closely aligns with the standard block width in Manhattan (264 ft) and the block length (900 ft) is divisible by 300 ft. By using cells of 300 ft in length, location-specific features can be captured more accurately, allowing for a detailed street-by-street resolution that enhances risk analysis. After cleaning the cells that do not contain any data points (e.g., around lakes), the grid map with complete data is shown in Figure 1.

3.2. Empirical Bayes Method

The Empirical Bayes (EB) method was introduced to estimate crash frequency. It combines the observed crash frequency from real-world data and the expected crash frequency predicted by the Safety Performance Function (SPF) [6]. In our study, we gathered observed crash data for different grids including all types of crashes. Then, a grid-based SPF was developed using the classic Negative Binomial (NB) model. The grid-based SPF allowed us to calculate an expected crash frequency for a particular grid. Finally, we applied the EB method. Instead of relying solely on observed crashes or expected crashes as predicted by the SPF, the EB method combines these two sources of information using the following Equations (1) and (2):

$$N_{eb} = \omega \times N_{spf} + (1 - \omega) \times N_{ob} \quad (1)$$

$$\omega = \frac{1}{1 + N_{spf}/\varphi} \quad (2)$$

where N_{eb} is the estimated crash frequency in each grid using the EB method, N_{spf} is the estimated crash frequency in each grid by the SPF, N_{ob} is the observed crash frequency in each grid, ω is the weight, and φ is the dispersion parameter of the SPF model.

The effectiveness of the EB method relies on the quality and reliability of the observed crash data and the accuracy of the SPF model. The integration with near-miss data is intended to build a more accurate grid-based SPF model.



Figure 1. Grid generation in the study area.

3.3. Spatial Analysis

Global Moran's I and local Moran's I were utilized to analyze the spatial autocorrelation of the estimated crash frequency. Global Moran's I and local Moran's I [37,38] spatial dependence tests were applied to measure spatial autocorrelation for estimated crash frequency [39]. The global Moran's I test [38] is widely used to measure how related the values of a variable are based on the locations and the values of their neighbors. The Local Indicators of Spatial Association (LISA) test assumes that global Moran's I is a summation of individual cross-products.

The global Moran's I statistic [37] assesses whether the value of one variable at one location is associated with its value at neighboring locations. The z-score of Moran's I (Z_I) and pseudo p -value [40] obtained from the permutation test are used to assess the significance of Moran's I. Z_I can be computed as:

$$Z_I = (I - E[I]) / SD[I] \quad (3)$$

where $E[I]$ is the expectation of I and $SD[I]$ is the standard deviation of I . A positive Z_I indicates that the observation distribution is spatially clustered [41] and a pseudo p -value less than 0.05 confirms that I is statistically significant at the confidence level of 95% [42]. More details about global Moran's I practice can be found in previous studies [34,41,43].

LISA is often used to capture local spatial patterns, and evaluates whether the patterns are clustered, dispersed, or randomly distributed. Local Moran's I for the observation z_i, z_j in cell i, j with weight matrix w_{ij} and N observations [44] can be computed as:

$$I_i = \frac{z_i}{\left(\frac{\sum_i z_i^2}{N}\right)} \sum_j w_{ij} z_j \quad (4)$$

4. Data Preparation

4.1. Near-Misses

The near-miss data were extracted from Mobileye collision warning events that were reported by vehicles equipped with Mobileye Advanced driver assistance system (ADAS) solution. A Collision Warning (CW) event is an alert generated when a Mobileye-equipped vehicle is on a trajectory to collide with another vehicle, pedestrian, or bicyclist in its path. Mobileye's Collision Warning system is based on the estimated Time to Collision (TTC), a function of velocity and distance. There are three types of collision warnings (see Figure 2) reported: Forward Collision Warning (FCW), Pedestrian Collision Warning (PCW), and Bicyclist Collision Warning (BCW). An FCW indicates a potential vehicle-to-vehicle collision, detected up to 80 m ahead and active for speeds between 1 km/h and 200 km/h. The TTC threshold for an FCW is triggered at 2.7 s. Both BCWs and PCWs involve potential collisions with bicyclists and pedestrians, respectively, detected up to 28 m ahead and active for speeds between 1 km/h and 50 km/h. The TTC threshold for these warnings is 2 s.

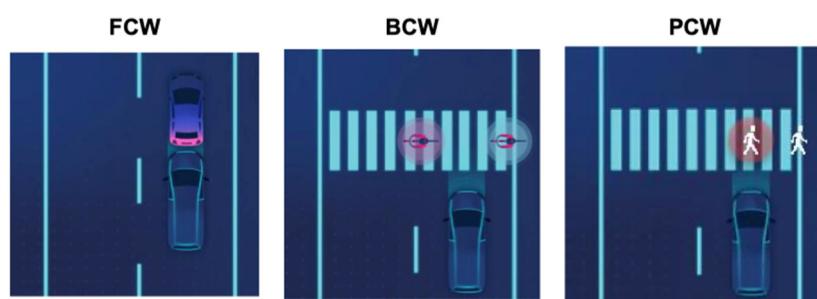


Figure 2. The concepts of FCWs, BCWs, and PCWs (Source: Mobileye).

The TTC threshold is crucial for defining near-misses. Since the near-miss data were provided by Mobileye, and the TTC thresholds were predetermined in the data source, we cannot modify them. FCWs, BCWs, and PCWs in the Mobileye data originate from collision warning systems. These systems typically use a TTC threshold of 2 to 3 s, allowing drivers enough time to react and take preventive action while minimizing unnecessary warnings. This threshold is supported by multiple studies [45,46] and guidelines [47,48]. In the literature, various TTC thresholds have been used to identify conflicts or near-misses, such as 1.5 s [49–51], 2.3 s [20], 2.5 s [52], and 3.5 s [53]. Although there is no unified TTC threshold, most studies use a threshold in the range of 1.5–3.5 s. Thus, the TTC thresholds in the source data fall within an acceptable range.

There are two types of Mobileye-equipped vehicles. The first type is called Mobileye 8 Connect (ME8), fleet vehicles with ME8 technology. All collision warnings generated from the ME8 vehicles are collected. The second type is called Original Equipment Manufacturer (OEM), consumer vehicles with OEM Mobileye technology. A subset of collision warnings is observed from OEM vehicles, and the default collection rate is set to gathering information from 1 driver per road segment per hour.

ME8 near-miss data used in this project cover the period from 5 July 2022, to 31 December 2022. OEM near-miss data cover the period from 16 August 2022, to 31 December 2022. During this timeframe, a total of 59,277 (ME8) and 2559 (OEM) warnings were generated, respectively (Figure 3), offering invaluable insights into potential collision events. In ME8 near-miss data, the vast majority of these warnings were FCWs (58,210 events), representing 98.2% of all warnings. PCWs and BCWs comprised a smaller portion of the dataset, with 711 (1.2%) and 356 (0.6%) warnings, respectively.

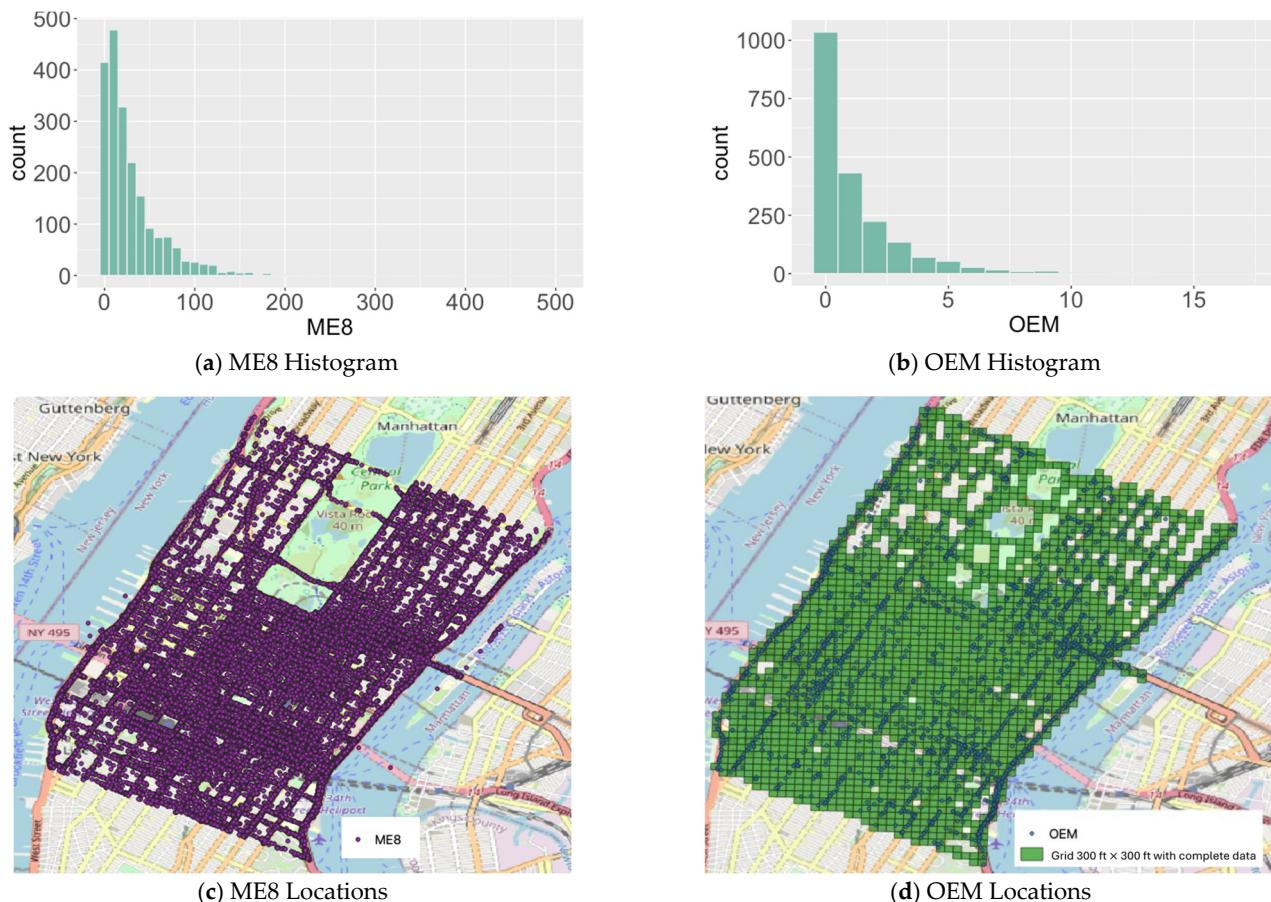


Figure 3. ME8 and OEM collision warning locations and distributions.

Similarly, OEM near-miss data consist of 91.8% FCWs (2005 events), 4.0% PCWs (88 events), and 4.2% BCWs (91 events). For each grid cell, the total count of the three collision warning types (FCW, PCW, and BCW) was calculated to represent the number of near-misses. This calculation was performed using the Spatial Join tool in ArcGIS.

4.2. Crash Data

The historical motor vehicle crash data for this analysis were obtained from Open NYC (<https://data.cityofnewyork.us/Public-Safety/Motor-Vehicle-Collisions-Crashes/h9gi-nx95> (accessed on 1 July 2023)). To maintain consistency with the near-miss data, the crash data were filtered to include incidents occurring between 5 July 2022 and 31 December 2022. Furthermore, any crash data entries lacking coordinate information were removed from the dataset. The crash data were then aggregated to each grid cell.

4.3. Other Data

4.3.1. Traffic Exposure Data

Annual Average Daily Traffic (AADT) and Vehicle Miles Traveled (VMT) data were used as approximations for traffic exposure. The most recent AADT data (2021) were acquired from the NYSDOT Traffic Data View website (<https://www.dot.ny.gov/tdv> (accessed on 28 June 2023)). As AADT data are a link-based feature of road networks, we calculated the VMT for each grid cell using the method proposed by the literature [35].

4.3.2. Road Network and Transport Facility

The road network data for New York City were obtained from [Data.gov](#). Subsequently, the total road length and the total number of intersections (road nodes) within each grid cell were calculated. To compute the total road length, the roads were divided at the grid

cell boundaries. We also aggregated the total counts of intersections as input. Moreover, a binary highway index was established for each grid cell based on the presence (highway index = 1) or absence (highway index = 0) of urban highways.

The total number of bus stops and subway stations per cell were also used as features of safety risk. The two latest shapefiles, showing NYC bus stops and subway stations in November 2020, were acquired from the website of the Newman Library of Baruch College, CUNY [54].

4.3.3. Land Use

Land use data were acquired from the NYC Department of City Planning (NYC DCP) Map PLUTO version 2022, v3.1 (<https://www.nyc.gov/site/planning/data-maps/open-data/dwn-pluto-mappluto.page> (accessed on 28 June 2023)). Each building class was assigned to the most appropriate land use category. We derived four land use category groups for this study, including residential area (R), commercial area (C), mixed residential and commercial area (Mixed R & C), and open space area (Figure 4). Land use data, represented as polygon features, were utilized to calculate the area percentage of each of the four land use category groups within the grid cells.

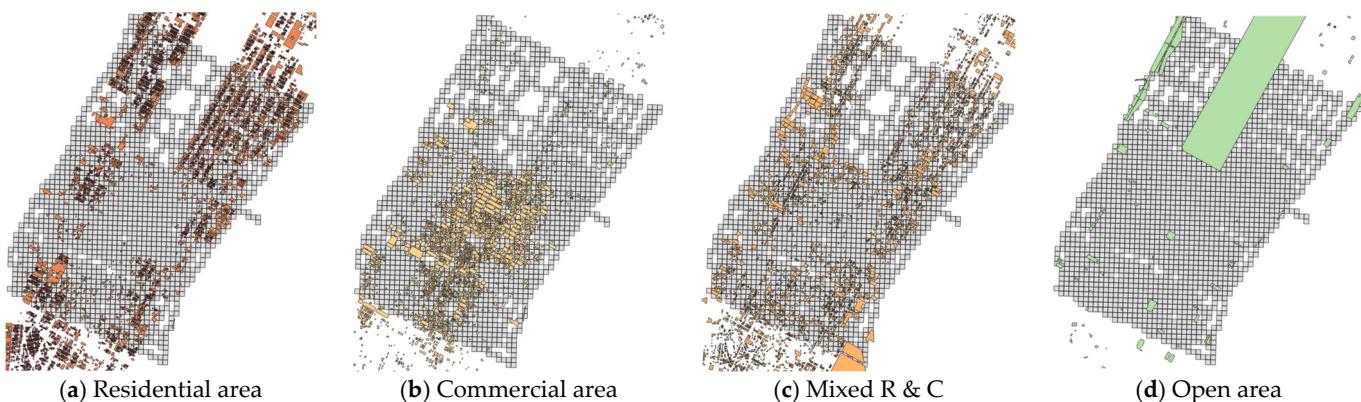


Figure 4. The land use group spatial distribution in the study area.

4.3.4. Population Density

Population data were obtained from the U.S. Census Bureau (<https://data.census.gov/> (accessed on 19 June 2024)), which offers various aggregation levels, such as census tracts, census tract block groups, and census tract blocks. The most recent data (2020 P1 Race table), aggregated at the census tract block level, were used for calculating the population within each grid cell. The calculation method for determining the population in each grid cell assumes a uniform distribution of the population within the region. The population within a grid cell originating from a specific census block was then calculated by multiplying the total population by the area ratio of the grid cell.

4.4. Variable Summary

The summary of the input variables used for safety analysis, including their mean, standard deviation (S.D.), median, minimum (Min), and maximum (Max) values, are provided in Table 1. The safety estimation was based on these variables, ensuring a comprehensive understanding of the potential factors influencing safety risk within the study area.

Table 1. Input variable summary.

Variable	Mean	S.D.	Median	Min	Max
Crash_tot	1.43	2.02	1	0	15
CW_ME8	30.87	38.30	19	0	504
CW_OEM	1.24	2.01	0	0	17
Intersection	0.93	1.40	1	0	18
Subway_station	0.026	0.17	0	0	2
Bus_stop	0.34	0.65	0	0	4
Road_length (ft)	407.97	176.82	355.22	3.64	1032.05
VMT (mi × veh)	1259	1869	723	0	15,777
Highway	0.08	0.27	0	0	1
Res_r	15%	20%	5%	0%	98%
Comm_r	16%	23%	2%	0%	87%
Open_r	9%	26%	0%	0%	100%
Mix_rc_r	14%	16%	8%	0%	91%
Population	259	226	227	0	1367

Note: Crash_tot is the total crash count in a grid, Res_r is the residential area rate in a grid, Comm_r is the commercial area rate in a grid, Open_r is the open area rate in a grid, Mix_rc_r is the mixed residential area and commercial area in a grid.

5. Results and Discussion

5.1. Correlation Analysis

We first examine the correlation between the variables using a correlation matrix, which is a statistical technique used to evaluate the relationship between a pair of variables in a dataset. As shown in Figure 5, the correlation coefficients of the total crash counts and other independent variables range from -0.15 to 0.43 . Specifically, the correlation coefficient of crash count and ME8 near-misses (CW_ME8) is 0.43 , which is the highest among those correlation coefficients for crash count. This indicates that the crash count (both total crashes and injury and fatal crashes) and near-misses have a moderate positive linear relationship and near-misses may play an important part in crash prediction models.

5.2. Crash Prediction Model

To provide a safety performance function that is necessary for the EB method, a Negative Binomial (NB) model [55,56] was employed. Firstly, we applied the correlation matrix (Figure 5) and Variance Inflation Factor (VIF) to identify highly correlated variables. Upon examining the correlations between variables, we observed that CW_OEM is highly correlated with both CW_ME8 and VMT. Given that CW_ME8 has a larger sample size and higher correlation coefficients with Crash_count, and considering that the sampling method of CW_OEM may be biased due to the data being gathered from one driver per road segment per hour, we decided to exclude CW_OEM from the model development. Another variable exhibiting high correlation coefficients is population, which is strongly correlated with Res_r and Mix_rc_r. Therefore, we opted to remove the population from subsequent steps. VMT and Highway index demonstrate a high correlation with one another. In consideration of the VIF results, where only the VIFs of VMT and Highway index exceed 2.5, we decided to exclude both variables to avoid multicollinearity issues.

Based on the variable selection process, we utilized the preliminarily selected variables to construct an NB model. The StepAIC function in R was employed to identify the optimal variables for input. The StepAIC function removed the Comm_r variable, and apart from the insignificant variables including Subway_station and mix_rc_r, all other variables were

found to be significant in the NB model. The resulting best NB model is presented in Table 2.

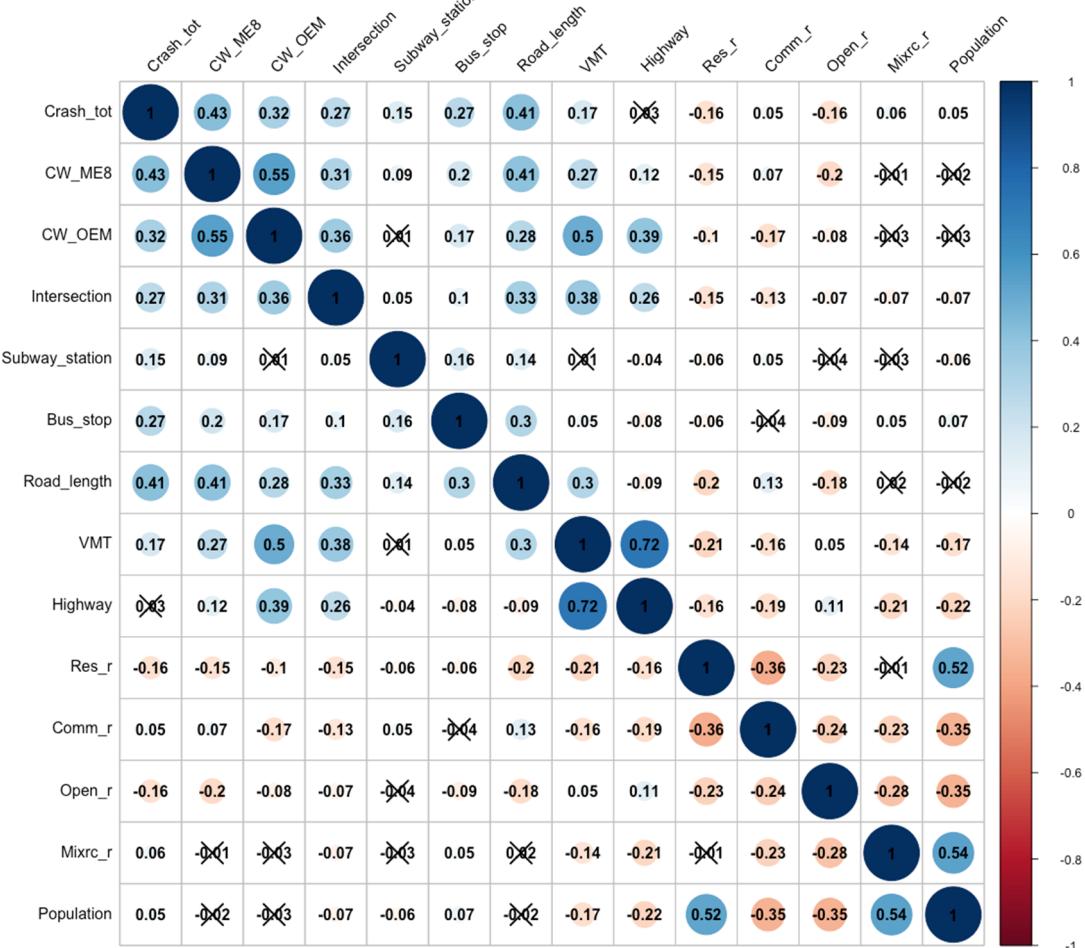


Figure 5. The correlation matrix of the safety-related variables. Note: a black cross on a number means the correlation coefficient is insignificant. Crash_tot is the total crash count in a grid.

Table 2. The estimated results of the best NB model.

	Estimate	Marginal Effects	Std. Error	z-Value	Pr(> z)
Intercept	-0.88		0.09	-9.389	0.00 ** ^a
CW_ME8	0.01	0.0074	0.00	10.694	0.00 **
Intersection	0.12	0.1322	0.02	6.893	0.00 **
Subway_station	0.20	0.2225	0.14	1.506	0.132
Bus_stop	0.19	0.2062	0.04	4.878	0.00 **
Road_length	0.01	0.0065	0.00	10.152	0.00 **
Res_r	-0.73	-0.8001	0.17	-4.419	0.00 **
Open_r	-1.04	-1.1295	0.16	-6.358	0.00 **
Mixrc_r	0.28	0.3002	0.17	1.587	0.112
Ψ^b	1.53				
Std. err	0.12				
AIC	6045.2				
2 × log-likelihood	-6025.2				

Note: ^(a) ** denotes significance at the 5% level. ^(b) Ψ is the dispersion parameter for the fitted NB model.

The marginal effect of near-misses, as represented by CW_ME8, is 0.0074. This implies that in our study area, the near-miss-to-crash rate is approximately 135:1. It is noteworthy that these near-miss data were sampled using Mobileye devices. To further refine the

total near-miss-to-crash rate, we leverage two statistics: that 10% of cars have the Advanced Driver Assistance System (ADAS) installed [57], and Mobileye holds a 69% market share in ADAS [58]. From this, we estimate that 6.9% of vehicles in New York City are equipped with Mobileye 8-connected devices. This, in turn, allows us to calculate the total near-miss-to-crash ratio (vehicle-to-vehicle near-miss: $1 \text{ km/h} < \text{speed} < 200 \text{ km/h}$, $\text{TTC} \leq 2.7 \text{ s}$, vehicle-to-bicycle or -pedestrian near-miss: $1 \text{ km/h} < \text{speed} < 50 \text{ km/h}$, $\text{TTC} \leq 2.0 \text{ s}$) in NYC as roughly 1957:1. Previous studies have provided empirical values for the near-miss/conflict-to-crash ratio. For instance, Gettman, et al. [59] suggested a ratio of approximately 20,000 conflicts to 1 actual crash, though the method for identifying conflicts was unspecified. A more recent study suggested a ratio of 14 conflicts (defined by a minimum TTC of less than 3.5 s) to 1 crash at regular signalized intersections [53]. While our results vary from the previous literature, they offer a reference near-miss-to-crash ratio for a dense urban environment using Mobileye-defined near-misses.

Our analysis also identified three significant network and facility-related variables: the number of intersections, the number of bus stops, and the road length. An increase in the number of intersections results in a rise in estimated crash frequency by 0.13. This increasing trend aligns with previous research [60,61]. This finding is reasonable given that intersections are spaces shared among various road users, potentially leading to higher crash exposure. Similarly, the presence of bus stops, where traffic disturbances and pedestrian conflicts are likely, raises the estimated crash frequency. An increase of one bus stop per grid enhances crash frequency by approximately 0.21, a finding trend in line with prior research [60]. The length of the road is positively associated with crash occurrences. An increase in road length by 100 ft augments the estimated crash count by 0.65, potentially due to increased exposure. In urban areas, longer roads often correlate with higher traffic volumes, a finding supported by the previous literature [62].

Conversely, two land use variables, the residential land use rate and the open space land use rate, negatively impact the estimated crash frequency. Residential areas usually have lower traffic volumes, possibly due to lower speed limits and more traffic calming measures, leading to fewer crashes [63]. Similarly, open spaces, with fewer vehicles and intersections, contribute to reduced crash rates, a finding supported by previous research [64]. However, these relationships can depend on numerous factors, such as road design and maintenance, and the behavior of drivers and pedestrians [65].

5.3. Spatial Analysis

5.3.1. Spatial Distribution

We visualized the EB estimated crash frequency map in Figure 6a. Since the range of the crash frequency estimated by EB (CF-EB) varies from 0.13 to 13.25, we divided it into 14 intervals, assigning a unique color to each interval. To facilitate comparison, we created a parallel visualization in Figure 6b, which ranks the CF, assigning a lower rank to grids with lower CF. The ranks are then divided into ten intervals (each encompassing 10% of the grids), with each interval given a unique color.

Figure 6 reveals two prominent patterns. First, grids encompassing linkage roads of bridges and tunnels exhibit a higher CF-EB, a finding that echoes previous research [64]. This is likely due to the complexity of the road network at these locations, with more frequent merging and diverging points leading to more frequent lane changes by drivers. This increased complexity can augment the potential for driver errors, contributing to higher safety risks in these areas. Additionally, significant traffic disruptions at these locations may heighten the likelihood of conflicts among motor vehicles, pedestrians, and bicyclists [64]. Second, high-risk grids predominantly line the avenues from 5th Ave to 8th Ave. These avenues are characterized by high pedestrian activity due to the concentration of commercial establishments, public transit access points, and other urban amenities. The heightened pedestrian activity, coupled with vehicular traffic, can escalate the potential for conflicts and safety risks.

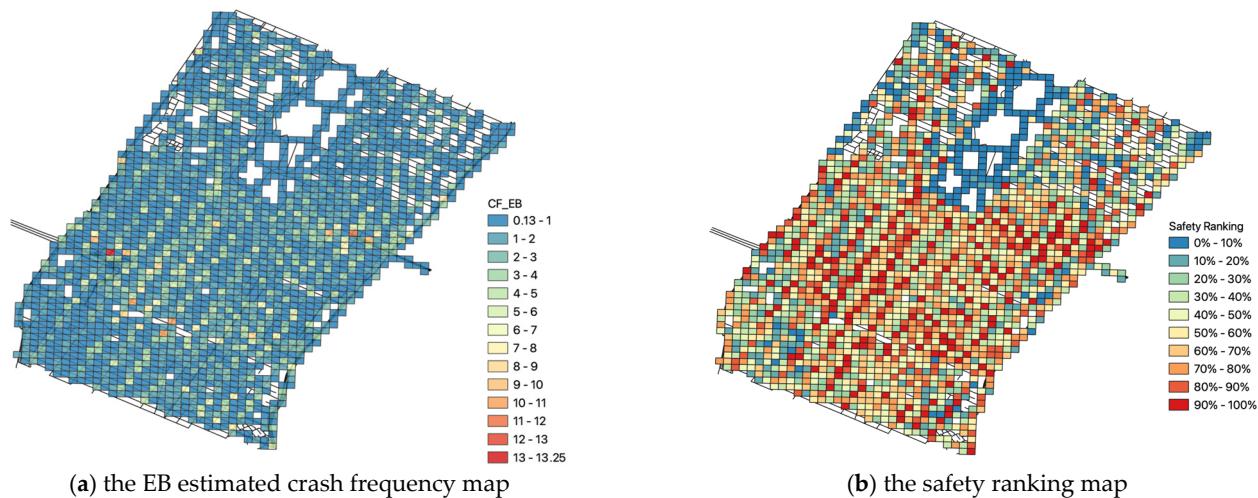


Figure 6. The spatial map of grid-based EB-estimated crash frequency.

5.3.2. Spatial Autocorrelation

We formulated the null hypothesis as the presence of complete spatial randomness. To test this, we conducted the Moran's I test using two k -nearest-neighbors weight matrices, with $k = 4$ and $k = 8$. With all the p -values ($k = 4, p = 0.003$; $k = 8, p = 0.001$) being statistically significant and positive z-scores ($k = 4, Z_i = 3.2037$; $k = 8, Z_i = 6.6389$) obtained using different weight matrices, we reject the null hypothesis. The results show a statistically significant spatial correlation, implying that the spatial distribution of high or low values of CF-EB clusters more spatially than it would under random spatial processes.

Figure 7 presents the univariate Local Spatial Autocorrelation results, called LISA cluster maps, for CF-EB using different numbers of neighbors ($k = 8$ and $k = 4$). The patterns exhibited by Figure 7a,b are strikingly similar. Both univariate LISA cluster maps reveal more clusters with positive local spatial correlation (high–high and low–low) than those with negative local spatial correlation (high–low and low–high).

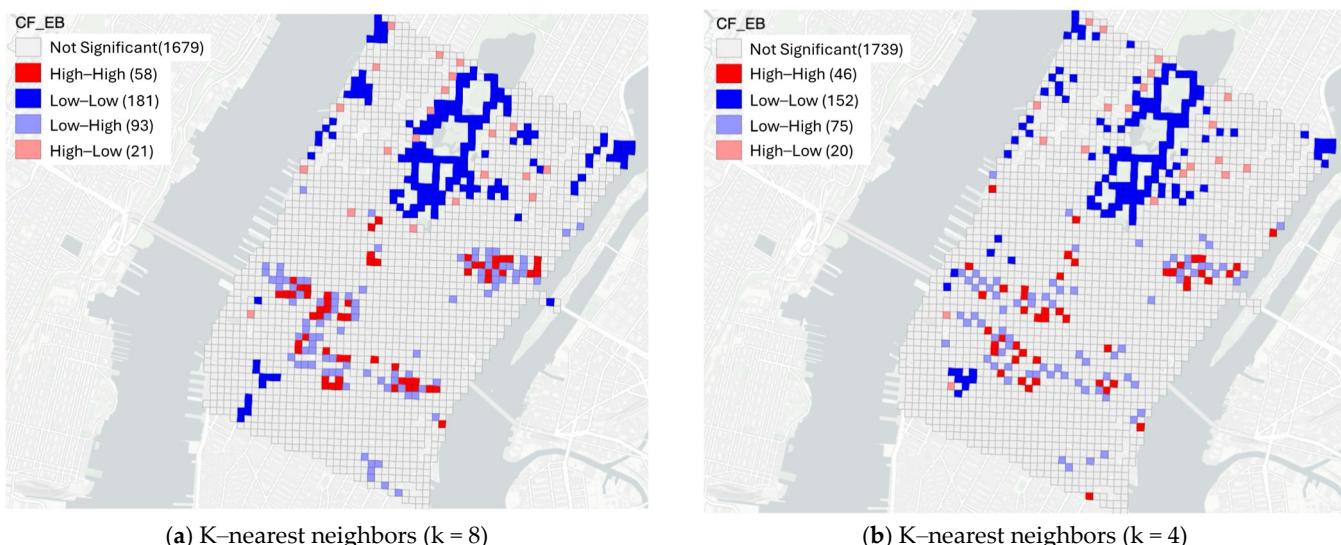


Figure 7. Univariate LISA cluster maps.

Locations with a positive local spatial correlation can be understood as high–high or low–low. In both univariate LISA cluster maps ($k = 4$ and $k = 8$), the low–low clusters are chiefly located around Central Park and the Chelsea area, representing low-risk locales. These low-risk areas could be the product of numerous factors such as well-structured road

infrastructure, effective traffic management, or lower traffic volumes. Large pedestrian zones around Central Park or lower vehicle speed limits might also contribute to reduced crash rates. High–high locales, on the other hand, predominantly encompass areas near the Queensboro Bridge linkage road and Midtown East, between 34th Street and 42nd Street. This could result from high traffic volume, intricate road systems, elevated vehicle speeds, or insufficient traffic control measures. The close proximity to the Queensboro Bridge, a significant transport route, may lead to increased traffic volumes and, consequently, more accidents.

In terms of negative local spatial correlation, both LISA cluster maps display fewer high–low locations but a significant number of low–high locations. High–low clusters primarily surround low–low clusters, such as the Central Park area. This arrangement suggests a sudden transition from low crash frequency regions to higher crash frequency regions, which could be caused by abrupt changes in road conditions, traffic volume, or other influential factors. Meanwhile, low–high clusters are predominantly situated near some high–high locations, such as the Queensboro Bridge linkage road area and Midtown East, between 34th Street and 42nd Street. Despite being surrounded by high-risk areas, these relatively safer grids might possess certain characteristics or interventions, such as lower speed limits, or well-designed intersections, that render them less susceptible to crashes.

5.3.3. Observation and Prediction Difference

The EB method relies on two crash values: the observed value and the predicted value. Analyzing the difference between these observed and predicted values can provide significant insights. Locations with large observation and prediction differences often warrant additional attention because they may either experience a high frequency of accidents or fewer accidents than expected [6,7]. We visualized this observation and prediction difference (OPD) in Figure 8. Figure 8a employs high opacity in the QGIS layer rendering parameters to present each grid's color, while Figure 8b uses low opacity and a burn blending mode to highlight grids with large absolute OPDs and to reveal the road network within each grid.

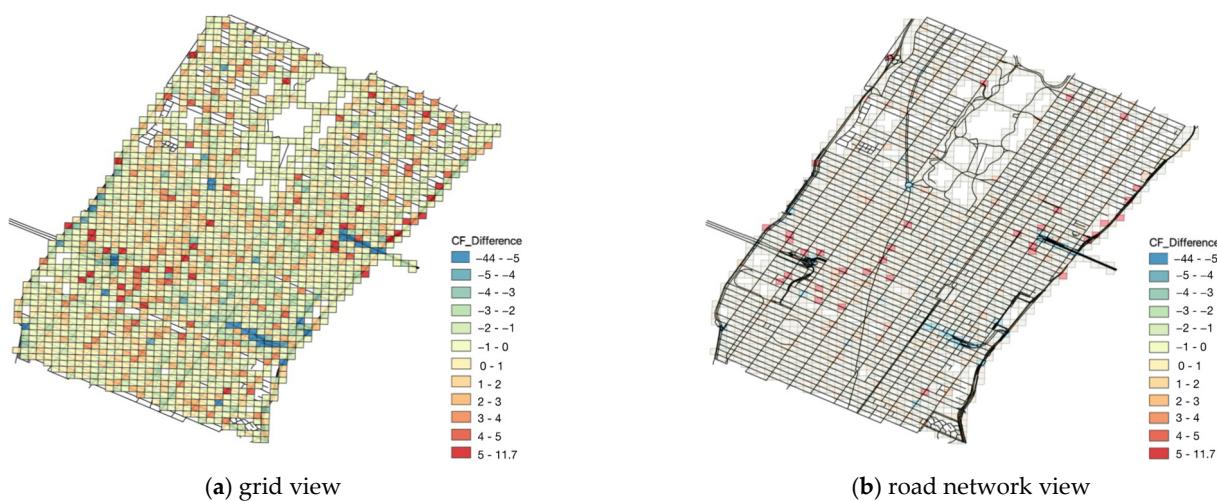


Figure 8. Observation and prediction difference between observation and model prediction.

In our study, OPD ranges from -44 to 11.7 . Notably, most grids exhibit small OPDs, demonstrating the efficacy of the SPF model across most grids. However, there are some grids with large absolute OPDs. Grids with OPD greater than 5 are colored dark red, while grids with OPD less than -5 are colored dark blue.

Dark red grids represent areas with a higher-than-expected crash frequency. As evident from Figure 8a, these grids are predominantly located near two tunnels, in the

vicinity of Penn Station, and along the highway. For the grids situated on the highway, the road network is characterized by risk factors such as curves [66] and interchanges [67]. For other grids, it is possible that high-risk intersections inside them which need further studies. On the other hand, the dark blue grids as seen in Figure 8b, which depict locations with fewer accidents than predicted, are situated on or near the linkage roads of bridges and tunnels (Lincoln Tunnel, Queens Midtown Tunnel, and Queensboro Bridge), or covering a roundabout (Columbus Circle). The road networks within these grids are often complex and with large road lengths. Such patterns may lead to the NB model's inadequacy, which is mainly fitted by grids with a simple road network, in accurately predicting the crash frequency for these abnormal grids.

6. Conclusions

This study integrates city-wide near-misses to estimate urban traffic safety and analyzes its spatial patterns. Grid-based aggregation method, the Empirical Bayes (EB) approach, and spatial analysis tools including global Moran's I and local Moran's I were applied to conduct this study.

The findings suggest that near-misses have the highest correlation with crash frequency among all the variables. Other variables such as the number of intersections, number of bus stops, road length, residential land use rate, and open-space land use rate significantly influence crash frequency in New York City. The near-miss-to-crash ratio, estimated to be 1957:1 for the study urban area, can serve as a potential benchmark for other urban environments. The analysis of network and facility-related variables revealed that a higher number of intersections and bus stops, and longer road length, all contribute to an increased crash frequency. Residential and open-space land use rates, on the other hand, were negatively correlated with crash frequency, likely due to lower traffic volumes, fewer complex intersections, and more traffic calming measures in these areas. Spatial analysis highlighted areas with high and low crash frequencies, revealing potential risk hotspots, such as linkage roads of bridges and tunnels, and avenues with high pedestrian activity. It was interesting to observe negative local spatial correlations in crash frequencies, pointing to significant variations in safety risks within short distances. The observation and prediction difference map further identified grids with unexpectedly high or low crash frequencies, offering valuable insights for targeted interventions.

Based on the findings, several detailed policy recommendations can help increase traffic safety in urban environments. Enhancing near-miss data collection and analysis by implementing city-wide near-miss reporting systems using mobile applications and connected vehicle technologies may provide comprehensive data for predictive models, identifying high-risk locations before crashes occur. Improving intersection safety remains one of the top priority actions in urban areas. Optimizing bus stop locations by conducting safety audits and relocating high-risk stops is. Encouraging residential and open space land use in urban planning may naturally improve safety. Targeted safety interventions in identified high-risk areas, such as linkage roads of bridges and tunnels, and high pedestrian activity avenues, can address specific safety concerns.

This research has potential implications for urban transportation safety policy and practice. It highlights the value of incorporating near-miss data collection and analysis into safety estimation. Understanding crash prediction and spatial distribution patterns can help city planners and policy makers devise and implement location-specific strategies for improving road safety.

Although the study offers valuable insights, certain limitations and future studies are presented. Although we can filter Vulnerable Road User (VRU)-related data from both crash and near-miss data, there are too few near-misses for VRUs to conduct statistical analysis. It would be interesting to check the estimated traffic safety for the two most common VRUs, pedestrians and cyclists, in the future. The variables used in this study, although comprehensive, do not account for all potential factors that may influence crash frequency, for example, curve and interchange. Further, the NB model, though it works well for most

grids, may not accurately represent grids with complex road networks. Future research could refine and expand upon this work by incorporating more variables, considering other modeling techniques, and extending the research to other urban environments. Moreover, studying driver behavior in near-miss events is crucial in the future, as it may significantly impact the relationship between near-misses and actual crashes. In addition, some conclusions could be further supported by additional data and analysis. For example, integrating hospital records related to traffic crashes could provide a more comprehensive understanding of injury severity and outcomes. Additionally, detailed behavioral data on driver and pedestrian actions during near-misses would enhance the analysis of the near-miss-to-crash relationship. Improving data diversity and quality can also be a key target for future studies.

Author Contributions: The authors confirm their contribution to the paper as follows: study conception and design: C.X., J.G., F.Z. and K.O.; data collection: C.X. and J.G.; analysis and interpretation of results: C.X., J.G., F.Z. and K.O.; draft manuscript preparation: C.X., J.G., F.Z. and K.O. All authors have read and agreed to the published version of the manuscript.

Funding: The work is funded by the C2SMART Center under Grant Number 69A3351747124 from the U.S. Department of Transportation's University Transportation Centers Program.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to a confidentiality agreement signed during the experiment.

Acknowledgments: The contents of this paper reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. The work in this paper is partially funded by the C2SMART University Transportation Center through a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof. The authors also gratefully acknowledge Mobileye for providing the near-miss data.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- WHO. Road Traffic Injuries. Available online: <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries> (accessed on 25 July 2023).
- ASIRT. ROAD SAFETY FACTS. Available online: <https://www.asirt.org/safe-travel/road-safety-facts/> (accessed on 29 May 2023).
- CHEKPEDS. NYC Crash Mapper. Available online: <https://crashmapper.org/> (accessed on 20 April 2022).
- FHWA. FHWA Strategic Plan 2022–2026. Available online: <https://highways.dot.gov/about/fhwa-strategic-plan-2022-2026> (accessed on 29 May 2023).
- NYC Vision Zero. What It Is. Available online: <https://www.nyc.gov/content/visionzero/pages/what-it-is> (accessed on 29 May 2023).
- Hauer, E.; Harwood, D.W.; Council, F.M.; Griffith, M.S. Estimating safety by the empirical Bayes method: A tutorial. *Transp. Res.* **2002**, *1784*, 126–131. [[CrossRef](#)]
- Huang, H.; Chin, H.C.; Haque, M.M. Empirical evaluation of alternative approaches in identifying crash hot spots: Naive ranking, empirical bayes, full bayes methods. *Transp. Res.* **2009**, *2103*, 32–41. [[CrossRef](#)]
- Sanders, R.L. Perceived traffic risk for cyclists: The impact of near miss and collision experiences. *Accid. Anal. Prev.* **2015**, *75*, 26–34. [[CrossRef](#)] [[PubMed](#)]
- Arun, A.; Haque, M.M.; Washington, S.; Sayed, T.; Mannering, F. A systematic review of traffic conflict-based safety measures with a focus on application context. *Anal. Methods Accid. Res.* **2021**, *32*, 100185. [[CrossRef](#)]
- Arun, A.; Haque, M.M.; Bhaskar, A.; Washington, S.; Sayed, T. A systematic mapping review of surrogate safety assessment using traffic conflict techniques. *Accid. Anal. Prev.* **2021**, *153*, 106016. [[CrossRef](#)] [[PubMed](#)]
- Park, J.-I.; Kim, S.; Kim, J.-K. Exploring spatial associations between near-miss and police-reported crashes: The Heinrich's law in traffic safety. *Transp. Res. Interdiscip. Perspect.* **2023**, *19*, 100830. [[CrossRef](#)]
- Wu, K.-F.; Jovanis, P.P. Crashes and crash-surrogate events: Exploratory modeling with naturalistic driving data. *Accid. Anal. Prev.* **2012**, *45*, 507–516. [[CrossRef](#)] [[PubMed](#)]

13. Tarko, A.; Davis, G.; Saunier, N.; Sayed, T.; Washington, S. *Surrogate Measures of Safety*; White Paper; Transportation Research Board: Washington, DC, USA, 2009.
14. Amundsen, F.; Hyden, C. *Proceedings of First Workshop on Traffic Conflicts*; TTI: Oslo, Norway; LTH Lund: Lund, Sweden, 1977; Volume 78.
15. Shahdah, U.; Saccomanno, F.; Persaud, B. Integrated traffic conflict model for estimating crash modification factors. *Accid. Anal. Prev.* **2014**, *71*, 228–235. [[CrossRef](#)]
16. Hyden, C.; Linderholm, L.; Swedish traffic-conflicts technique, T. *International Calibration Study of Traffic Conflict Techniques*; Springer: Berlin/Heidelberg, Germany, 1984; pp. 133–139.
17. Dijkstra, A. Assessing the safety of routes in a regional network. *Transp. Res. Part C Emerg. Technol.* **2013**, *32*, 103–115.
18. Machiani, S.G.; Abbas, M. Safety surrogate histograms (SSH): A novel real-time safety assessment of dilemma zone related conflicts at signalized intersections. *Accid. Anal. Prev.* **2016**, *96*, 361–370. [[CrossRef](#)]
19. Zheng, L.; Ismail, K.; Meng, X. Investigating the heterogeneity of postencroachment time thresholds determined by peak over threshold approach. *Transp. Res. Rec.* **2016**, *2601*, 17–23. [[CrossRef](#)]
20. Xie, K.; Yang, D.; Ozbay, K.; Yang, H. Use of real-world connected vehicle data in identifying high-risk locations based on a new surrogate safety measure. *Accid. Anal. Prev.* **2019**, *125*, 311–319. [[CrossRef](#)] [[PubMed](#)]
21. Guo, H.; Xie, K.; Keyvan-Ekbatani, M. Modeling driver’s evasive behavior during safety-critical lane changes: Two-dimensional time-to-collision and deep reinforcement learning. *Accid. Anal. Prev.* **2023**, *186*, 107063. [[CrossRef](#)] [[PubMed](#)]
22. Yang, D.; Xie, K.; Ozbay, K.; Yang, H.; Budnick, N. Modeling of time-dependent safety performance using anonymized and aggregated smartphone-based dangerous driving event data. *Accid. Anal. Prev.* **2019**, *132*, 105286. [[CrossRef](#)] [[PubMed](#)]
23. Lyu, N.; Cao, Y.; Wu, C.; Xu, J.; Xie, L. The effect of gender, occupation and experience on behavior while driving on a freeway deceleration lane based on field operational test data. *Accid. Anal. Prev.* **2018**, *121*, 82–93. [[CrossRef](#)]
24. Papadoulis, A.; Quddus, M.; Imprailou, M. Evaluating the safety impact of connected and autonomous vehicles on motorways. *Accid. Anal. Prev.* **2019**, *124*, 12–22. [[CrossRef](#)]
25. Cai, P.; Wang, S.; Wang, H.; Liu, M. Carl-lead: Lidar-based end-to-end autonomous driving with contrastive deep reinforcement learning. *arXiv* **2021**, arXiv:2109.08473.
26. Ortiz, F.M.; Sammarco, M.; Detyniecki, M.; Costa, L.H.M. Road traffic safety assessment in self-driving vehicles based on time-to-collision with motion orientation. *Accid. Anal. Prev.* **2023**, *191*, 107172. [[CrossRef](#)]
27. Anisha, A.M.; Abdel-Aty, M.; Abdelraouf, A.; Islam, Z.; Zheng, O. Automated vehicle to vehicle conflict analysis at signalized intersections by camera and LiDAR sensor fusion. *Transp. Res. Rec.* **2023**, *2677*, 117–132. [[CrossRef](#)]
28. Gecchele, G.; Orsini, F.; Gastaldi, M.; Rossi, R. Freeway rear-end collision risk estimation with extreme value theory approach. A case study. *Transp. Res. Procedia* **2019**, *37*, 195–202. [[CrossRef](#)]
29. Wu, J.; Xu, H.; Zheng, Y.; Tian, Z. A novel method of vehicle-pedestrian near-crash identification with roadside LiDAR data. *Accid. Anal. Prev.* **2018**, *121*, 238–249. [[CrossRef](#)]
30. Xu, C.; Ozbay, K.; Liu, H.; Xie, K.; Yang, D. Exploring the impact of truck traffic on road segment-based severe crash proportion using extensive weigh-in-motion data. *Saf. Sci.* **2023**, *166*, 106261. [[CrossRef](#)]
31. Ali, Y.; Haque, M.M.; Mannering, F. A Bayesian generalised extreme value model to estimate real-time pedestrian crash risks at signalised intersections using artificial intelligence-based video analytics. *Anal. Methods Accid. Res.* **2023**, *38*, 100264. [[CrossRef](#)]
32. Fu, C.; Sayed, T. Identification of adequate sample size for conflict-based crash risk evaluation: An investigation using Bayesian hierarchical extreme value theory models. *Anal. Methods Accid. Res.* **2023**, *39*, 100281. [[CrossRef](#)]
33. AASHTO. *Highway Safety Manual*; AASHTO: Washington, DC, USA, 2010.
34. Gao, J.; Xie, K.; Ozbay, K. Exploring the spatial dependence and selection bias of double parking citations data. *Transp. Res. Rec.* **2018**, *2672*, 159–169. [[CrossRef](#)]
35. Xie, K.; Ozbay, K.; Kurkcu, A.; Yang, H. Analysis of traffic crashes involving pedestrians using big data: Investigation of contributing factors and identification of hotspots. *Risk Anal.* **2017**, *37*, 1459–1476. [[CrossRef](#)] [[PubMed](#)]
36. Xie, K.; Ozbay, K.; Zhu, Y.; Yang, H. Evacuation zone modeling under climate change: A data-driven method. *J. Infrastruct. Syst.* **2017**, *23*, 04017013. [[CrossRef](#)]
37. Moran, P.A.P. Notes on Continuous Stochastic Phenomena. *Biometrika* **1950**, *37*, 17–23. [[CrossRef](#)]
38. Moran, P.A. The interpretation of statistical maps. *J. R. Stat. Society. Ser. B* **1948**, *10*, 243–251.
39. Tiefelsdorf, M. The saddlepoint approximation of Moran’s I’s and local Moran’s li’s reference distributions and their numerical evaluation. *Geogr. Anal.* **2002**, *34*, 187–206.
40. Anselin, L.; Syabri, I.; Kho, Y. GeoDa: An introduction to spatial data analysis. *Geogr. Anal.* **2006**, *38*, 5–22. [[CrossRef](#)]
41. Xie, K.; Ozbay, K.; Yang, H. Spatial analysis of highway incident durations in the context of Hurricane Sandy. *Accid. Anal. Prev.* **2015**, *74*, 77–86. [[CrossRef](#)] [[PubMed](#)]
42. Goodchild, M.F. *Spatial Autocorrelation*; Geo Books: Kerala, India, 1986.
43. Xie, K.; Wang, X.; Ozbay, K.; Yang, H. Crash frequency modeling for signalized intersections in a high-density urban road network. *Anal. Methods Accid. Res.* **2014**, *2*, 39–51. [[CrossRef](#)]
44. Luc, A. Local Indicators of Spatial Association—LISA. *Geogr. Anal.* **1995**, *27*, 93–115. [[CrossRef](#)]
45. Wang, M.; Liao, Y.; Lyckvi, S.L.; Chen, F. How drivers respond to visual vs. auditory information in advisory traffic information systems. *Behav. Inf. Technol.* **2020**, *39*, 1308–1319. [[CrossRef](#)]

46. Deveaux, D.; Higuchi, T.; Uçar, S.; Wang, C.-H.; Härrí, J.; Altintas, O. Extraction of Risk Knowledge from Time to Collision Variation in Roundabouts. In Proceedings of the 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), Indianapolis, IN, USA, 19–22 September 2021; pp. 3665–3672.
47. Wilson, T.B.; Butler, W.; McGehee, D.V.; Dingus, T.A. *Forward-Looking Collision Warning System Performance Guidelines*; SAE International: Warrendale, PA, USA, 1997; pp. 701–725.
48. Vasudevan, M.; O’Hara, J.; Townsend, H.; Asare, S.; Muhammad, S.; Ozbay, K.; Yang, D.; Gao, J.; Kurkcu, A.; Zuo, F. *Algorithms to Convert Basic Safety Messages into Traffic Measures*; National Academy of Sciences: Washington, DC, USA, 2022.
49. Jeong, E.; Oh, C.; Lee, G.; Cho, H. Safety impacts of intervehicle warning information systems for moving hazards in connected vehicle environments. *Transp. Res. Rec.* **2014**, *2424*, 11–19. [CrossRef]
50. Genders, W.; Razavi, S.N. Impact of connected vehicle on work zone network safety through dynamic route guidance. *J. Comput. Civ. Eng.* **2016**, *30*, 04015020. [CrossRef]
51. Kondyli, A.; Schrock, S.D.; Tousif, F. *Evaluation of Near-Miss Crashes Using a Video-Based Tool*; Kansas Department of Transportation. Bureau of Research: Topeka, KS, USA, 2023.
52. So, J.J.; Dedes, G.; Park, B.B.; HosseinyAlamdary, S.; Grejner-Brzezinska, D. Development and evaluation of an enhanced surrogate safety assessment framework. *Transp. Res. Part C Emerg. Technol.* **2015**, *50*, 51–67. [CrossRef]
53. Borsos, A.; Farah, H.; Laureshyn, A.; Hagenzieker, M. Are collision and crossing course surrogate safety indicators transferable? A probability based approach using extreme value theory. *Accid. Anal. Prev.* **2020**, *143*, 105517. [CrossRef]
54. GIS Lab at Newman Library of Baruch College, CUNY, NYC Mass Transit Spatial Layers. 2020. Available online: <https://www.baruch.cuny.edu/confluence/display/geoportal/NYC+Mass+Transit+Spatial+Layers> (accessed on 25 July 2023).
55. Gardner, W.; Mulvey, E.P.; Shaw, E.C. Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychol. Bull.* **1995**, *118*, 392. [CrossRef]
56. Zou, Y.; Zhang, Y.; Lord, D. Application of finite mixture of negative binomial regression models with varying weight parameters for vehicle crash data analysis. *Accid. Anal. Prev.* **2013**, *50*, 1042–1051. [CrossRef] [PubMed]
57. Canalys. Huge Opportunity as only 10% of the 1 Billion Cars in Use Have ADAS Features. Available online: <https://www.canalys.com/newsroom/huge-opportunity-as-only-10-of-the-1-billion-cars-in-use-have-adas-features> (accessed on 29 July 2023).
58. Krause, R. Can Mobileye Extend Its Auto Software Dominance Into Driverless Cars? Available online: <https://www.investors.com/research/the-new-america/mobileye-stock-will-its-auto-software-dominance-extend-to-driverless-cars/> (accessed on 29 July 2023).
59. Gettman, D.; Pu, L.; Sayed, T.; Shelby, S.G.; Energy, S. *Surrogate Safety Assessment Model and Validation*; Turner-Fairbank Highway Research Center: McLean, VA, USA, 2008.
60. Kim, K.; Pant, P.; Yamashita, E. Accidents and accessibility: Measuring influences of demographic and land use variables in Honolulu, Hawaii. *Transp. Res. Rec.* **2010**, *2147*, 9–17. [CrossRef]
61. Siddiqui, C.; Abdel-Aty, M.; Choi, K. Macroscopic spatial analysis of pedestrian and bicycle crashes. *Accid. Anal. Prev.* **2012**, *45*, 382–391. [CrossRef] [PubMed]
62. Han, C.; Huang, H.; Lee, J.; Wang, J. Investigating varying effect of road-level factors on crash frequency across regions: A Bayesian hierarchical random parameter modeling approach. *Anal. Methods Accid. Res.* **2018**, *20*, 81–91. [CrossRef]
63. Lee, G.; Joo, S.; Oh, C.; Choi, K. An evaluation framework for traffic calming measures in residential areas. *Transp. Res. Part D Transp. Environ.* **2013**, *25*, 68–76. [CrossRef]
64. Xie, K.; Ozbay, K.; Yang, D.; Xu, C.; Yang, H. Modeling bicycle crash costs using big data: A grid-cell-based Tobit model with random parameters. *J. Transp. Geogr.* **2021**, *91*, 102953. [CrossRef]
65. Luo, Y.; Liu, Y.; Xing, L.; Wang, N.; Rao, L. Road safety evaluation framework for accessing park green space using active travel. *Front. Environ. Sci.* **2022**, *10*, 864966. [CrossRef]
66. Bartin, B.; Ozbay, K.; Xu, C. Extracting horizontal curvature data from GIS maps: Clustering method. *Transp. Res. Rec.* **2019**, *2673*, 264–275. [CrossRef]
67. Yao, Y.; Zhao, X.; Li, J.; Ma, J.; Zhang, Y. Traffic safety analysis at interchange exits using the surrogate measure of aggressive driving behavior and speed variation. *J. Transp. Saf. Secur.* **2023**, *15*, 515–540. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.