Digital Transportation
and Safety

ARTICLE

Open Access

https://doi.org/10.48130/dts-0024-0016

Digital Transportation and Safety **2024**, 3(4): 184−190

Spatial insights into micro-mobility safety: establishing optimal buffers for scooter crash predictions

Boniphace Kutela[1*], Meshack P. Mihayo[2], Emmanuel Kidando[2], Tumlumbe Juliana Chengula[3] and Sia M. Lyimo[4]

[1] *Texas A&M Transportation Institute, 701 N Post Oak Ln #430, Houston, TX 77024, USA*
[2] *Department of Civil and Environmental Engineering, Cleveland State University, 2121 Euclid Avenue, Cleveland, OH 44115, USA*
[3] *South Carolina State University, 300 College Avenue, Orangeburg, SC 29117, USA*
[4] *Progressive AE, 1811 4 Mile Rd NE, Grand Rapids, MI 49525, USA*
* Corresponding author, E-mail: b-kutela@tti.tamu.edu

Abstract

Establishing comparison events/crashes is among the key challenges in safety analysis. This study proposes a spatial consideration for predicting scooter crashes using Utah's five years of crash data. It involves creating buffers ranging from 5 to 250 ft from the point of the scooter crash to obtain comparison crashes. The appropriate variables were selected based on the literature and engineering judgment. The Binary Logistic Regression was then applied to determine the appropriate buffer based on the consistency in the direction and magnitude of the impact of predictor variables. Results indicate that three variables, the junction type, lighting condition, and weather condition, are susceptible to changes in the direction of impact. Moreover, the study findings reveal that as the buffer distance increases, the magnitude of the impact of the variables decreases. Based on the results, a buffer of less than 50 ft is deemed appropriate for various analyses due to consistency in direction and the magnitude of impact. Further, the study findings show that intersections, dark-lighted conditions, summer season, and right-turning movements are more likely to be associated with scooter crashes. These findings can be crucial to transportation agencies and practitioners in improving the safety of scooter riders.

Keywords: Scooter safety; Micro-mobility; Spatial analysis; Optimal buffer

Citation: Kutela B, Mihayo MP, Kidando E, Chengula TJ, Lyimo SM. 2024. Spatial insights into micro-mobility safety: establishing optimal buffers for scooter crash predictions. *Digital Transportation and Safety* 3(4): 184−190 https://doi.org/10.48130/dts-0024-0016

Introduction

Among the micro-mobility devices used in the United States, electric scooters (E-scooters) have been reported to have the highest proportion of injuries and fatalities, with 68 deaths from 2017 through 2021[1]. This can be attributed to the difficulty of controlling the scooter compared to other micro-mobility devices because of the high instability due to their short wheelbase, small wheels, and relatively high center of gravity[2,3]. Previous studies found that at least half of the scooter victims sustained severe injuries or even deaths[4]. Further, at least half of scooter riders incurred injuries to the face/head[5]. Transportation agencies and manufacturers have focused on improving riders' safety. Accurate prediction of scooter crashes is imperative to understanding the key contributing factors associated with scooter crashes.

Although various studies have been performed to understand the risk factors associated with scooter crashes, establishing comparison cases/crashes to scooter crashes has been challenging. Scooter crashes usually represent a relatively small portion of the crash data, thus introducing class imbalance. Most previous studies compared scooter crashes against all other crashes in the database[6,7]. However, this approach may not provide appropriate results as it compares locations where scooter crashes occurred against locations with no scooter crashes. Some studies have utilized machine learning-based approaches like SMOTE to balance the data to determine comparison crashes[8]. However, such an approach selects data randomly and may include crashes from locations without scooter crashes. A buffer approach has been utilized to determine crashes that occurred in homogenous characteristics[9]. The buffering approach is a relatively better option as it captures crashes at the given location with similar characteristics—for instance, a study by Avelar et al. utilized a buffer of 250 ft to define intersection-related crashes[10]. Although the 250 ft buffer has predominantly been used, it is unclear how useful/practical it is for specific types of crashes, especially those involving micro-mobility or vulnerable road users. These crashes will likely cover a small portion, which may be less than 250 ft. Further, several changes might be observed in a 250 ft distance. For instance, the impact of lighting conditions can vary at this distance. A scooter rider at 250 ft is relatively less likely to be illuminated by the lighting at that distance. Thus, a better understanding of the space consideration for scooter crashes is necessary.

This study presents spatial considerations for the prediction of scooter crashes. It utilizes space-constrained data to predict the likelihood of scooter crash occurrence compared to other crashes. Various researchers have employed crash comparison analysis[9,11−13]. Researchers have proven that using 'space-constrained data' results in better estimates of the impacts and attributes of crashes rather than utilizing the full dataset[9]. To better understand the key contributing factors associated with scooter crashes compared to other crashes, this study employed a space-constrained approach to predict scooter crashes. Findings from this study can be utilized to develop appropriate countermeasures to improve scooter safety.

Literature review

E-scooter-related crashes have raised several health and safety concerns among the public and agencies. Previous studies have explored the risks of scooter riding[14,15], attributes of E-scooter-related crashes, and injury crashes of scooter riders[15,16]. Key factors associated with scooter crashes include infrastructure, behavioral, and environmental[4,5,14−16].

Previous studies show that transitioning from one riding surface to another increases the likelihood of scooter crashes. Further, using off-road for riding a scooter increases crash risk by 24 times[15]. Other studies have explored scooter crash locations and collision types. Yang et al. found that intersections, sidewalks, and arterial roads/streets are critical locations for scooter crashes[4]. The behaviors of scooter riders and other road users—who mostly interact with scooters—significantly influence scooter safety. Yang et al. explored the need to wear helmets among scooter riders and found that 76.7% of scooter riders who did not wear helmets were severely injured or dead[4]. Various researchers have also studied alcohol usage[4,5]. It has been found that riding under the influence (alcohol usage) increases the severity of scooter crashes[5]. Moreover, riding during non-daylight hours increases the risk of scooter crashes by 5.52 times compared to riding during daylight hours. Also, scooter riding along a two-way directional traffic flow was associated with a 1.72 times higher risk of encountering crashes than riding in a one-directional traffic flow[15]. Regarding collision types, Yang et al. found that major collision types of scooters fall off and collide with other vehicles, whereby the proportion of 'hitting vehicle' was much higher than that of 'falling off'[4]. Additionally, the study investigated the effects of scooter riders' demographics on the crash severity, and findings show that female scooter riders are more vulnerable to severe injuries than male riders; however, statistics show that more male riders were deceased due to scooter crashes compared to female riders.

Various scooter data collection techniques have been utilized in previous studies depending on factors such as study objective, data accessibility, sampling strategy, the scope of the study, etc. For instance, to explore the safety risks of scooters, Ma et al.[14] and White et al.[15] performed naturalistic riding experiments to collect scooter-related data through sensors and cameras installed in the scooters. Data collected include scooter trips, critical events, experienced vibrations, speed changes, and proximity of nearby objects. Other studies have utilized police-reported crash data in their analysis. For instance, Shah et al. explored 52 scooter and 79 bicycle police-reported crashes to compare motor vehicle-related scooter crashes and bicycle crashes[6]. Also, Blackman & Haworth utilized police-reported crash data to compare the risk and severity of motorcycle, moped, and larger scooter crashes[7]. Moreover, Karpinski et al. extracted data on scooter fatalities involving motor vehicles from the FARS database to compare motor vehicle-involved scooter fatalities with other fatalities[17]. Yang et al. gathered scooter crash data from media news reports to describe patterns of crashes related to scooter riding[4].

To analyze scooter-related crashes, Yang et al. performed a descriptive and cross-tabulation analysis of the mined media news reports to analyze key attributes and their interactions concerning scooter crashes[4]. White et al. developed data mining algorithms to extract recorded data from the cameras and sensors installed in 50 scooters to quantify safety risk factors of scooters based on behavioral, environmental, and infrastructure aspects[15]. Ma et al. conducted a statistical analysis using R to program workflow for detecting scooter vibrations and speed variations[14]. Blackman & Haworth developed an ordered probit model using SPSS software to compare the impacts of scooter crashes among motorcycles, mopeds, and larger scooters[7]. Tian et al. employed the Negative Binomial Regression approach to explore the risk factors of scooter-related crashes[16]. Moreover, to compare motor vehicle-involved scooter and bicycle crashes, Shah et al. employed two analysis methods: descriptive analysis and crash typology[6]. Further, the study conducted a t-statistic test to evaluate the difference in means of scooter crashes and bicycle crashes. In addition, Karpinski et al. performed a statistical test of proportions to compare scooter fatalities with other fatalities[17]. Thus, it has been observed that most of the previous scooter-related studies have utilized a 'full crash dataset' from a certain database, hospital records, police records, controlled or naturalistic data collection techniques, and simulation to analyze scooter-related crashes.

In summary, various studies have been performed to understand factors associated with scooter crashes and their associated severity. However, most scooter-related studies have not established comparison crashes in their analyses. Comparison crashes are a crucial technique, especially when it is necessary to understand the key contributing factors associated with crash occurrences of a certain type of crash compared to other crashes. Moreover, unlike the full dataset approach, comparison crashes enable better estimates of the key factors associated with crash occurrences[9]. Therefore, this study utilizes spatial considerations to extract comparison crashes to predict the likelihood of scooter crashes.

Material and methodology

This section presents the approach used to predict scooter crashes compared to other crashes. The study employed binary logistic regression analysis to evaluate the likelihood of scooter crashes for different pre-defined buffer distances.

Data description

This study utilized five years (2018 to 2022) of statewide crash data from Utah to predict scooter crashes based on spatial considerations[18]. The dataset contained all types of crashes, including scooter and non-scooter crashes. It also contained important attributes like crash severity, junction type, lighting condition, weather conditions, driver age, etc.

First, the scooter crashes were identified using the vehicle type variable and crash narrative. A total of 260 scooter crashes were identified. Six buffers, 5, 10, 50, 100, 150, and 250 ft, were created for each scooter crash to determine the comparison of crashes to scooter crashes. For each buffer, corresponding crashes and associated features/attributes of interest were extracted.

Further, based on the literature review, only important attributes that impact scooter crashes were selected from the original crash dataset for further analysis. Six categorical variables, junction type, lighting condition, weather condition, vehicle maneuver, driver age, and season of the year, and one continuous variable, traffic volume, were selected for further analysis.

Statistical modeling approach

In this study, scooter crashes were compared to other crashes. As such, two response variables were involved: (i) scooter-related crashes and, (ii) non-scooter-related crashes. These binary outcomes were modeled using a binary logit model due to its simplicity and interpretability regarding the odds ratio. Further, the binary logit model assumes that the observations are not from repeated measurements and that the independent variables have little or no multicollinearity[19,20]. The binary logistic regression is represented using a Bernoulli probability function whereby the response variable Y_i has two binary outcomes (1 or 0). The probability of the scooter crash $P(Y_i = 1)$ can be expressed as an inverse logistic function of a vector of explanatory variables X_i as depicted in Eqn (1) below:

$$P(Y_i = 1) = \frac{1}{1 + e^{-z}} \tag{1}$$

$$z = \hat{\beta}_0 + \hat{\beta}_i x_i \tag{2}$$

$$Odds\ Ratio\,(OR) = \frac{P}{1 - P} \tag{3}$$

whereby, $\hat{\beta}_i$ represent variable coefficients to be estimated while $\hat{\beta}_0$ is a constant term. The maximum likelihood method was used to estimate the variable coefficients. Moreover, in the binary outcomes, the 0 value represents all other comparison crashes, while 1 represents scooter crashes.

The interpretation of the model was based on the odds ratio (OR) since the variable coefficients cannot provide a straightforward meaning for the logistic regression model. Given the event of interest, which in this study is the scooter crash, a certain variable will be associated with the increase in the probability of the event of interest if its OR is greater than one. However, if the OR is less than one, the variable is associated with a decrease in the probability of the event of interest. If the OR of the variable is equal to one, it implies that the variable does not influence the event of interest[21].

Descriptive analysis

Table 1 shows variables and their descriptive statistics for each buffer. Generally, throughout the buffers, about 47.69% of scooter crashes occurred at T-intersections, 86.15% occurred in lighting conditions other than dark-lighted, 93.08% occurred in other weather conditions other than cloudy, 50.77% occurred when vehicles turned right, 89.23% occurred in age groups other than older drivers, and 56.96% occurred during the fall season.

For other crashes, the number of observations differs across buffers. However, the trend in each buffer for other crashes has remained the same as what is presented for the scooter crashes. That means higher proportions of other crashes have been observed to occur at T-intersections, in weather conditions other than cloudy, when vehicles are involved in other movements, and age groups other than older drivers. For example, for a 5 ft buffer, 82.35% of comparison crashes have been observed to occur at T-intersections. 82.35% occurred in lighting conditions other than dark-lighted, 89.08% occurred in weather conditions other than cloudy, 22.5% occurred when vehicles were involved in other movements, 78.57% occurred in age groups other than older, 37.18% occurred during the summer season.

Model results and discussion

Table 2 presents the logistic regression results for scooter crashes for the six buffer distances of 5, 10, 50, 100, 150, and 250 ft. From Table 2, the prediction accuracy for the buffer-specific model increased with buffer distances from 10 to 250 ft. Also, the direction (sign) and magnitude of the impact of the variable coefficients changed across the different buffer distances. Below is a discussion of the observed changes in prediction accuracy and the direction and magnitude of the variable coefficients.

Model performance

Results in Table 2 show a clear trend of changes in the prediction capabilities of the logistic regression model as the buffer size increases. Although not uniform, the prediction accuracy generally increased with increased buffer distance. The lowest prediction accuracies of 87.5% and 84.3% corresponded to the smallest 5-ft and 10-ft buffer distances, respectively. On the other hand, the prediction accuracy increased monotonically with the buffer distance. The prediction accuracy comprises two components: the true positive and the true negative. In this case, the true positive values score represents the actual scooter crashes, while the true negative is non-scooter crashes. Due to data imbalance, most observations were in the non-event class of interest. Thus, it signifies the increase in non-scooter crashes. However, if the objective is to predict scooter crashes, the model with a high prediction accuracy would not be approximated because the overall prediction accuracy of the model represents the non-event class due to data imbalance. A similar scenario in the context of crash prediction has been observed in previous studies whereby sensitivity, specificity, and precision were employed as the supplemental model performance measures to consider both event and non-event classes[22]. Considering the models'' prediction capabilities, the 5 ft buffer is deemed appropriate for predicting scooter crashes as it has a relatively high prediction accuracy, sensitivity, and specificity score.

Direction and magnitude of variable coefficients

To understand an appropriate buffer distance for scooter crash prediction, it is also important to consider the possibility of changes in the direction of impact. The results in Table 2 indicate that the direction of impact of some variables changed, as indicated by the positive and negative signs of the estimates/coefficients. In particular, three variables, the junction type, lighting condition, and weather condition, showed a change in the direction of the impact across the buffers. Among the variables, junction type is statistically significant at a 95% confidence level for the 5 ft buffer. Per the results, the model that used a 5 ft buffer revealed that non-intersection locations are more likely to be associated with scooter crashes. On the other hand, a 250 ft buffer model indicates that non-intersection locations are less likely to be associated with scooter crashes. Such observation underscores the importance of space constraints when performing comparative analysis. The 250 ft buffer distance might include other features not associated with the scooter crash location, thus showing no influence on scooter crashes. The junction type and lighting conditions are spatial features, and when the buffer increases, there is a high likelihood of including features unrelated to the scooter crash. Other attributes, such as dark-lighted and cloudy weather conditions, were not statistically significant at a 95%

Table 1. Descriptive summary of the variables.

Variable name	Category	5-ft buffer				10-ft buffer				50-ft buffer			
		Other crashes		Scooter crashes		Other crashes		Scooter crashes		Other crashes		Scooter crashes	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
Junction type	4-Leg intersection	196	67.8%	124	47.7%	372	81.8%	124	47.7%	2186	82.0%	124	47.7%
	T-Intersection	8	2.8%	45	17.3%	20	4.4%	45	17.3%	101	3.79%	45	17.3%
	Non-intersection	15	5.2%	40	15.4%	30	6.6%	40	15.4%	190	7.13%	40	15.4%
	Others	19	6.6%	51	19.6%	33	7.3%	51	19.6%	189	7.09%	51	19.6%
Lighting condition	Daylight	178	61.6%	198	76.2%	337	74.1%	198	76.2%	1943	72.88%	198	76.2%
	Dark-lighted	42	14.5%	36	13.8%	91	20.0%	36	13.8%	574	21.53%	36	13.8%
	Others	18	6.2%	26	10.0%	27	5.9%	26	10.0%	149	5.59%	26	10.0%
Weather condition	Clear	186	64.4%	226	86.9%	370	81.3%	226	86.9%	2079	77.98%	226	86.9%
	Cloudy	26	9.0%	18	6.9%	46	10.1%	18	6.9%	375	14.07%	18	6.9%
	Others	26	9.0%	16	6.2%	39	8.6%	16	6.2%	212	7.95%	16	6.2%
Vehicle Maneuver	Straight ahead	4	1.4%	117	45.0%	9	2.0%	117	45.0%	58	2.18%	117	45.0%
	Stopped vs straight ahead	21	7.3%	6	2.3%	30	6.6%	6	2.3%	227	8.51%	6	2.3%
	Both straight	56	19.4%	5	1.9%	106	23.3%	5	1.9%	582	21.83%	5	1.9%
	Straight vs turning left	69	23.9%	5	1.9%	131	28.8%	5	1.9%	824	30.91%	5	1.9%
	Right turn	23	8.0%	67	25.8%	54	11.9%	67	25.8%	292	10.95%	67	25.8%
	Other movements	65	22.5%	60	23.1%	125	27.5%	60	23.1%	683	25.62%	60	23.1%
Driver age	Older driver - No	187	64.7%	232	89.2%	361	79.3%	232	89.2%	2203	82.63%	232	89.2%
	Older driver - Yes	51	17.6%	28	10.8%	94	20.7%	28	10.8%	463	17.37%	28	10.8%
Season of the Year	Winter	52	18.0%	21	8.1%	108	23.7%	21	8.1%	669	25.09%	21	8.1%
	Spring	58	20.1%	47	18.1%	103	22.6%	47	18.1%	593	22.24%	47	18.1%
	Summer	46	15.9%	90	34.6%	100	22.0%	90	34.6%	585	21.94%	90	34.6%
	Fall	82	28.4%	102	39.2%	144	31.6%	102	39.2%	819	30.72%	102	39.2%

Variable name	Category	100-ft buffer				150-ft buffer				250-ft buffer			
		Other crashes		Scooter crashes		Other crashes		Scooter crashes		Other crashes		Scooter crashes	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
Junction type	4-Leg intersection	3403	79.0%	124	47.7%	3849	74.5%	124	47.7%	4221	66.7%	124	47.7%
	T-Intersection	203	4.7%	45	17.3%	320	6.2%	45	17.3%	562	8.9%	45	17.3%
	Non-intersection	274	6.4%	40	15.4%	317	6.1%	40	15.4%	426	6.7%	40	15.4%
	Others	430	10.0%	51	19.6%	677	13.1%	51	19.6%	1119	17.7%	51	19.6%
Lighting condition	Daylight	3165	73.4%	198	76.2%	3775	73.1%	198	76.2%	4641	73.3%	198	76.2%
	Dark-lighted	908	21.1%	36	13.8%	1092	21.2%	36	13.8%	1307	20.7%	36	13.8%
	Others	237	5.5%	26	10.0%	296	5.7%	26	10.0%	380	6.0%	26	10.0%
Weather condition	Clear	3346	77.6%	226	86.9%	4001	77.5%	226	86.9%	4885	77.2%	226	86.9%
	Cloudy	627	14.5%	18	6.9%	730	14.1%	18	6.9%	902	14.3%	18	6.9%
	Others	337	7.8%	16	6.2%	432	8.4%	16	6.2%	541	8.5%	16	6.2%
Vehicle Maneuver	Straight ahead	129	3.0%	117	45.0%	175	3.4%	117	45.0%	245	3.9%	117	45.0%
	Stopped vs straight ahead	543	12.6%	6	2.3%	703	13.6%	6	2.3%	861	13.6%	6	2.3%
	Both straight	889	20.6%	5	1.9%	1029	19.9%	5	1.9%	1189	18.8%	5	1.9%
	Straight vs turning left	1049	24.3%	5	1.9%	1130	21.9%	5	1.9%	1301	20.6%	5	1.9%
	Right turn	456	10.6%	67	25.8%	511	9.9%	67	25.8%	612	9.7%	67	25.8%
	Other movements	1244	28.9%	60	23.1%	1615	31.3%	60	23.1%	2120	33.5%	60	23.1%
Driver age	Older driver - No	3604	83.6%	232	89.2%	4325	83.8%	232	89.2%	5313	84.0%	232	89.2%
	Older driver - Yes	706	16.4%	28	10.8%	838	16.2%	28	10.8%	1015	16.0%	28	10.8%
Season of the Year	Winter	1094	25.4%	21	8.1%	1332	25.8%	21	8.1%	1620	25.6%	21	8.1%
	Spring	920	21.3%	47	18.1%	1094	21.2%	47	18.1%	1353	21.4%	47	18.1%
	Summer	980	22.7%	90	34.6%	1168	22.6%	90	34.6%	1438	22.7%	90	34.6%
	Fall	1316	30.5%	102	39.2%	1569	30.4%	102	39.2%	1917	30.3%	102	39.2%

confidence interval for the 250 ft buffer. However, they indicated that the increased buffer distance may affect their results.

The changes in the directions of impact for the three variables (junction type, lighting condition, and weather condition) started at different buffers. For instance, that change starts to be noticed at a 150-ft buffer for non-intersection locations. These findings imply that there shouldn't be a single value/buffer for everything. The buffer should vary depending on the purpose of the analysis. For instance, a buffer of less than 50 ft can be used for lighting conditions assessment, while for intersection crashes, a buffer of less than 150 ft should be adopted. This is contrary to the previous studies that applied a buffer of 250 ft for safety analysis of intersections[10].

The increase in buffer distance was also associated with the changes in the magnitude of the impact of the variables. Overall, as the buffer distance increased, the magnitude of the impact decreased, as indicated by the decreasing magnitudes of coefficients and associated odd ratios. The impact of the increased buffer distance was more pronounced in the spatially controlled factors such as the lighting condition and

Table 2. Logistic regression model results.

		5-ft buffer			10-ft buffer			50-ft buffer		
		Estimate	OR	*p*-value	Estimate	OR	*p*-value	Estimate	OR	*p*-value
Intercept		**6.172**	**478.98**	**< 0.001**	**4.571**	**96.62**	**<0.001**	**2.774**	**16.02**	**0.002**
Junction type	*T-intersection*	**1.400**	**4.05**	**0.007**	**1.100**	**3.00**	**0.007**	**1.095**	**2.99**	**<0.001**
	Non-intersection	**1.176**	**3.24**	**0.020**	**0.956**	**2.60**	**0.020**	**0.845**	**2.33**	**0.003**
	Others	**2.450**	**11.59**	**< 0.001**	**1.940**	**6.96**	**< 0.001**	**1.756**	**5.79**	**<0.001**
Lighting condition	*Dark-lighted*	**1.460**	**4.31**	**0.001**	***0.624***	***1.87***	***0.070***	−0.016	0.98	0.948
	Others	0.885	2.42	0.115	**1.122**	**3.07**	**0.021**	0.483	1.62	0.147
Weather condition	*Cloudy*	0.020	1.02	0.972	0.628	1.87	0.191	−0.089	0.92	0.782
	Others	***−0.933***	***0.39***	***0.089***	−0.230	0.79	0.619	−0.147	0.86	0.685
Vehicle maneuver	*Stopped vs straight ahead*	**−6.223**	**0.002**	**< 0.001**	**−5.476**	**0.004**	**<0.001**	**−5.560**	**0.004**	**< 0.001**
	Both straight	**−5.134**	**0.01**	**< 0.001**	**−5.034**	**0.01**	**<0.001**	**−4.829**	**0.01**	**< 0.001**
	Straight vs turning left	**−6.080**	**0.002**	**< 0.001**	**−5.722**	**0.003**	**<0.001**	**−5.532**	**0.004**	**< 0.001**
	Turning right	**1.702**	**5.49**	**< 0.001**	**1.296**	**3.66**	**<0.001**	**0.947**	**2.58**	**< 0.001**
	Other movements	**−3.262**	**0.04**	**< 0.001**	**−3.100**	**0.05**	**<0.001**	**−2.834**	**0.06**	**< 0.001**
Traffic volume	*Log (AADT)*	**−1.123**	**0.33**	**0.001**	**−0.965**	**0.38**	**0.001**	**−0.895**	**0.41**	**< 0.001**
Driver age	*Older driver*	**−1.020**	**0.36**	**0.012**	**−0.732**	**0.48**	**0.035**	***−0.510***	***0.60***	***0.064***
Season of the year	*Spring*	0.519	1.68	0.357	0.712	2.04	0.114	***0.619***	***1.86***	***0.070***
	Summer	**1.640**	**5.15**	**0.003**	**1.695**	**5.45**	**< 0.001**	**1.450**	**4.26**	**< 0.001**
	Fall	**1.051**	**2.86**	**0.041**	**1.390**	**4.01**	**0.001**	**1.291**	**3.64**	**< 0.001**
Number of observations			429			641			2777	
AIC			331			475			922	
Prediction accuracy			87.5%			84.3%			92.3%	
Sensitivity score			85.5%			91.7%			97.4%	
Specificity score			89.8%			67.8%			55.9%	

		100-ft buffer			150-ft buffer			250-ft buffer		
		Estimate	OR	*p*-value	Estimate	OR	*p*-value	Estimate	OR	*p*-value
Intercept		**1.859**	**6.42**	**0.017**	**1.959**	**7.09**	**0.009**	**2.124**	**8.36**	**0.004**
Junction type	*T-intersection*	**1.234**	**3.44**	**< 0.001**	**1.242**	**3.46**	**< 0.001**	**0.943**	**2.57**	**< 0.001**
	Non-intersection	0.222	1.25	0.388	−0.086	0.92	0.729	**−0.493**	**0.61**	**0.040**
	Others	**1.424**	**4.15**	**< 0.001**	**1.157**	**3.18**	**< 0.001**	***0.465***	***1.59***	***0.053***
Lighting condition	*Dark-lighted*	−0.050	0.95	0.826	−0.084	0.92	0.704	−0.159	0.85	0.466
	Others	0.336	1.40	0.265	0.225	1.25	0.444	0.125	1.13	0.662
Weather condition	*Cloudy*	−0.248	0.78	0.420	−0.316	0.73	0.298	−0.386	0.68	0.195
	Others	−0.381	0.68	0.270	−0.365	0.69	0.276	−0.416	0.66	0.203
Vehicle maneuver	*Stopped vs straight ahead*	**−5.593**	**0.004**	**< 0.001**	**−5.684**	**0.003**	**< 0.001**	**−5.639**	**0.004**	**< 0.001**
	Both straight	**−4.508**	**0.01**	**< 0.001**	**−4.461**	**0.01**	**< 0.001**	**−4.418**	**0.01**	**< 0.001**
	Straight vs turning left	**−5.057**	**0.01**	**< 0.001**	**−4.985**	**0.01**	**< 0.001**	**−4.884**	**0.01**	**< 0.001**
	Turning right	**0.961**	**2.61**	**< 0.001**	**1.039**	**2.83**	**< 0.001**	**1.031**	**2.80**	**< 0.001**
	Other movements	**−2.548**	**0.08**	**< 0.001**	**−2.557**	**0.08**	**< 0.001**	**−2.533**	**0.08**	**< 0.001**
Traffic volume	*Log(AADT)*	**−0.823**	**0.44**	**< 0.001**	**−0.895**	**0.41**	**< 0.001**	**−0.928**	**0.40**	**< 0.001**
Driver age	*Older driver*	−0.372	0.69	0.149	−0.321	0.73	0.206	−0.387	0.68	0.120
Season of the year	*Spring*	**0.711**	**2.04**	**0.027**	**0.772**	**2.16**	**0.015**	**0.670**	**1.95**	**0.031**
	Summer	**1.309**	**3.70**	**< 0.001**	**1.396**	**4.04**	**< 0.001**	**1.301**	**3.67**	**< 0.001**
	Fall	**1.175**	**3.24**	**< 0.001**	**1.272**	**3.57**	**< 0.001**	**1.166**	**3.21**	**< 0.001**
Number of observations			4341			5109			6165	
AIC			1123			1188			1283	
Prediction accuracy			95.5%			96.0%			96.9%	
Sensitivity score			99.1%			99.4%			99.8%	
Specificity score			18.6%			11.9%			8.5%	

Bolded values are statistically significant at a 95% while ***bolded and italic*** are statistically significant at a 90% confidence interval.

intersection type. For instance, while the T-intersection was over four times more likely to be associated with scooter crashes for the 5 ft model, the magnitude was about 2.57 times for the 250 ft model. A similar trend was observed for non-intersection crashes. It was even worse for other junction types where the buffer increase from 5 to 250 ft was associated with a decrease in about 10 times odd ratios.

Other variables that significantly changed the magnitude of impact include right turn movements and the summer season. Turning movements can be associated with spatial factors, such as increased buffer distance, including some driveways where vehicles are likely to turn right. The remaining variable showed a compact nature irrespective of the increase in the buffer distance.

Based on the results from this study, due to the difference in the specificity scores from a 50 to 100 ft buffer. While the specificity scores for 5, 10, and 50 ft buffers were 89.8%, 67.8%, and 55.9%, respectively, the 100, 150, and 250 ft buffers had specificity scores of 18.6%, 11.9%, and 5.8%. Due to the drop in the specificity scores from the 50 to 100 ft buffer, a buffer less or equal to 50 ft may be deemed appropriate for various analyses. According to the results, intersection type, right turning movements, dark-lighted conditions, and summer seasons are the key predictors of scooter crashes. The group of other intersections, such as roundabouts, etc., was over 11 times more likely to be associated with scooter crashes. The T-intersections were also highly likely, as indicated by a 4.05 odds ratio. This implies that scooter crashes are generally more likely to occur at intersection locations than at non-intersection locations. Similar results were obtained from the study conducted by Shah et al., whereby a higher frequency of scooter-related crashes occurred at intersections[6]. However, the findings contradict those of Yang et al.[4]. According to their study, more scooter crashes were observed to occur on arterial streets rather than at intersections[4]. These contradictory results might be due to the nature of the analysis, as their study used reports from the news but did not compare scooter crashes with any other crash types.

Dark-lighted conditions had higher odds of crash occurrences (4.31 for a 5 ft buffer) than other lighting conditions within the same buffer. This finding is nearly similar to the one obtained by White et al., who found that riding during non-daylight conditions increases the risk of scooter crashes by 5.52 times compared to riding in other conditions[15]. However, the study did not consider spatial considerations in its analysis. The higher likelihood of scooter crashes in dark-lighted conditions is due to the low visibility associated with riding scooters during dark-lighted conditions, which tends to increase the likelihood of scooter crashes. The findings from this study contradict the ones by Shah et al., which found that daylight conditions are more associated with scooter crashes than other lighting conditions[6]. It should be noted that the study by Shah et al. was exploratory and did not compare scooter crashes with other crashes[6].

Summer seasons are associated with higher chances of scooter crashes. Scooter-related crashes have been observed to have a 5.15 times odds ratio during summer seasons. Previous studies support this finding. For instance, it has been found that the number of scooter crashes has peaked during the summer months compared to other months[4,5]. Also, Shah et al. observed higher crash rates of scooters in the summer[6]. This is because, during summer seasons, there is a higher utilization of e-mobility products like scooters compared to other seasons; as such, exposure to crashes also increases due to increased interactions between scooters and other vehicles[23].

Right-turning movements have been observed to influence scooter crashes with an odds ratio of 5.49 compared to other vehicle maneuvers. Based on the results in Table 2, the magnitude of impact for turning right movement has been observed to decrease considerably as the buffer distance increases. This can be explained by the fact that as buffer distance increases, more features might be incorporated into the analysis, which cannot be associated with the occurrence of the specific scooter crash. In addition, scooter riders interact much with motor vehicles, especially when the motor vehicles turn right; thus, such interactions contribute to the increase in the likelihood of scooter crashes. However, this is contrary to the findings by Shah et al., which found that most scooter crashes occurred when the vehicle was going straight while the scooter rider was crossing from the right side of the motorist[6]. In their study, Shah et al. applied descriptive analysis to 52 e-scooters and 79 bicycle police reported crashes in Nashville but did not compare their findings with any other crash[6]. Thus, their conclusion may hold only when comparing scooter crashes at various locations.

Conclusions and future studies

Understanding the key factors associated with scooter crashes compared to other crashes is imperative to improving the safety of scooters. Past studies reported that the majority of the scooter victims sustained severe injuries or even deaths. As such, accurate prediction of scooter crashes is highly important compared to other crashes. Unlike most previous studies utilizing the entire dataset or simulation methods, this study utilized a constrained dataset to predict scooter crashes based on predefined buffers. The study developed logistic regression models to predict the likelihood of scooter crashes to the predefined buffers: 5, 10, 50, 100, 150, and 250 ft.

This study presents the following significant observations based on the changes in prediction capabilities, changes in the direction of impact of the variables, and changes in the magnitude of the impact of the variables.

• As the buffer increases, different variables will predict the likelihood of a scooter crash differently. Therefore, it is important to select and use a suitable buffer size depending on the purpose of the specific analysis.

• As the buffer distance increases, the impact of different variables decreases. A significant decrease was observed in the following variables: junction type, lighting condition, right turning movement, and summer season – which are the key predictors of scooter crashes.

• A 50 ft or less buffer is deemed appropriate for various analyses. This is due to the better performance shown by models with less or equal to 50 ft buffer distance than models with greater or equal to 100 ft buffer distance using their three performance measures: prediction accuracy, sensitivity score, and specificity score.

This study investigated the impact of different buffers for crash analysis on predicting certain types of crashes, such as scooter crashes. It is evident from this study that the current and commonly used 250 ft buffer might not be a suitable criterion for scooter crash analysis. This study has shown that at the 250 ft buffer, the prediction accuracy and sensitivity of the model is high, but the specificity score is low, which implies that the model correctly predicts non-scooter crashes. Therefore, this research highlights the importance of selecting a proper buffer distance to analyze scooter crashes depending on the level of detail the intended analysis should offer. The research further provides future researchers with an opportunity to investigate the adaptability of a similar approach to other types of crashes that have few occurrences in nature but sustain serious injuries if they occur.

Various safety practitioners and transportation agencies can use the findings from this study to improve roadway safety in general and scooter/micromobility crashes specifically. The main application of this study is when developing crash modification factors (CMFs) for scooter/micro-mobility-related counter measures. Practitioners should use a buffer that is more appropriate to scooters/micromobility, contrary to the 250 ft buffer that is commonly used. This study suggests that a buffer of 50 ft should be used for safety analysis. The CMFs developed using wider buffers may produce unrealistic results. Various agencies rely on the CMFs to improve traffic safety. Thus, improving the approach to determine proper CMFs will benefit them.

Despite the findings, this study has the following limitations: first, the study used about 260 scooter crash data, which may be relatively small sample data. Future studies may consider expanding the sample size to include scooter crashes from different cities and jurisdictions. Second, the current study used regression analysis, which may not answer why the crash happened. Thus, future studies may consider other approaches, such as text analysis and related complex machine learning algorithms, to understand the variations of the reasons for crashes as the spatial consideration varies. Third, this study utilized scooter crash data from the State of Utah Department of Transportation, lacking multi-regional representation. Future studies may consider scooter crash data from different regions to ensure spatial distribution and adaptability. In addition, they should also consider adding population demographic-related dependent variables such as race, income, scooter typical use, etc., to capture as wide a population distribution as possible.

Author contributions

The authors confirm contribution to the paper as follows: study conception and design: Kutela B; data collection: Kutela B, Kidando E; analysis and interpretation of results: Kutela B, Mihayo MP, Kidando E; draft manuscript preparation: Kutela B, Mihayo MP, Kidando E, ChengulaT, Lyimo S. All authors reviewed the results and approved the final version of the manuscript.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Conflict of interest

The authors declare that they have no conflict of interest.

Dates

Received 24 March 2024; Revised 12 August 2024; Accepted 22 August 2024; Published online 27 December 2024

References

1. The U.S. Consumer Product Safety Commission (CPSC). 2022. *E-Scooter, E-Bike and Hoverboard Injuries and Deaths Are on the Rise; Celebrate National Fire Prevention Week with the Safe Use of Micromobility Products*. www.cpsc.gov/Newsroom/News-Releases/2023/E-Scooter-E-Bike-and-Hoverboard-Injuries-and-Deaths-Are-on-the-Rise-Celebrate-National-Fire-Prevention-Week-with-the-Safe-Use-of-Micromobility-Products (Accessed September 6, 2023)
2. Ringer K. 2019. *Design and Performance Assessment of a Novel Electric Scooter*. Thesis. University of California, Santa Cruz, USA. https://escholarship.org/uc/item/8wb872z6
3. Novotny A, Mollenhauer M, White E. 2023. *E-Scooter Design: Safety Measures for Next Gen Scooter*. Tech Report. Virginia, USA. https://rosap.ntl.bts.gov/view/dot/73277/dot_73277_DS1.pdf
4. Yang H, Ma Q, Wang Z, Cai Q, Xie K, et al. 2020. Safety of micromobility: analysis of E-Scooter crashes by mining news reports. *Accident Analysis & Prevention* 143:105608
5. Kleinertz H, Volk A, Dalos D, Rutkowski R, Frosch KH, et al. 2023. Risk factors and injury patterns of e-scooter associated injuries in Germany. *Scientific Reports* 13:706
6. Shah NR, Aryal S, Wen Y, Cherry CR. 2021. Comparison of motor vehicle-involved e-scooter and bicycle crashes using standardized crash typology. *Journal of Safety Research* 77:217–28
7. Blackman RA, Haworth NL. 2013. Comparison of moped, scooter and motorcycle crash risk and crash severity. *Accident Analysis & Prevention* 57:1–9
8. Sanjurjo-de-No A, Pérez-Zuriaga AM, García A. 2023. Analysis and prediction of injury severity in single micromobility crashes with Random Forest. *Heliyon* 9:e23062
9. Kutela B, Oscar C, Kidando E, Mihayo M. 2023. Seeing through eyes of Drivers: space consideration in investigating visibility of Vulnerable Road users involved in crashes from Driver's perspective. *Accident Analysis & Prevention* 192:107260
10. Avelar RE, Dixon KK, Escobar P. 2015. Evaluation of signalized-intersection crash screening methods based on distance from intersection. *Transportation Research Record: Journal of the Transportation Research Board* 2514:177–86
11. Novat N, Kidando E, Kutela B, Kitali AE. 2023. A comparative study of collision types between automated and conventional vehicles using Bayesian probabilistic inferences. *Journal of Safety Research* 84:251–60
12. Seacrist T, Douglas EC, Huang E, Megariotis J, Prabahar A, et al. 2018. Analysis of near crashes among teen, young adult, and experienced adult drivers using the SHRP2 naturalistic driving study. *Traffic Injury Prevention* 19:S89–S96
13. Chengula TJ, Kutela B, Novat N, Shita H, Kinero A, et al. 2024. Spatial instability of crash prediction models: a case of scooter crashes. *Machine Learning with Applications* 17:100574
14. Ma Q, Yang H, Mayhue A, Sun Y, Huang Z, et al. 2021. E-Scooter safety: the riding risk analysis based on mobile sensing data. *Accident Analysis & Prevention* 151:105954
15. White E, Guo F, Han S, Mollenhauer M, Broaddus A, et al. 2023. What factors contribute to e-scooter crashes: a first look using a naturalistic riding approach. *Journal of Safety Research* 85:182–91
16. Tian D, Ryan AD, Craig CM, Sievert K, Morris NL. 2022. Characteristics and risk factors for electric scooter-related crashes and injury crashes among scooter riders: a two-phase survey study. *International Journal of Environmental Research and Public Health* 19:10129
17. Karpinski E, Bayles E, Daigle L, Mantine D. 2023. Comparison of motor-vehicle involved e-scooter fatalities with other traffic fatalities. *Journal of Safety Research* 84:61–73
18. AASHTOWare. 2024. *AASHTOWare Safety*. https://udot.aashtoware-safety.com/signin (Accessed March 8, 2024)
19. Kutela B, Teng H. 2020. Evaluating the influential factors for push-button utilization at signalized midblock crosswalks. *Safety Science* 122:104533
20. Woodridge JM. 2012. *Introductory Economics: A modern Approach*. 5th Edition. Boston, MA, USA: Cengage Learning Publication.
21. UCLA. 2014. *Introduction to Generalized Linear Mixed Models*. Statistical Consulting Group. https://stats.oarc.ucla.edu/other/mult-pkg/introduction-to-generalized-linear-mixed-models/ (accessed November 29, 2023)
22. Zhou X, Lu P, Zheng Z, Tolliver D, Keramati A. 2020. Accident prediction accuracy assessment for highway-rail grade crossings using random forest algorithm compared with decision tree. *Reliability Engineering & System Safety* 200:106931
23. Miranda-Moreno LF, Nosal T, Schneider RJ, Proulx F. 2013. Classification of bicycle traffic patterns in five North American cities. *Transportation Research Record: Journal of the Transportation Research Board* 2339:68–79