Leveraging Social Media Data to Understand Spatial and Severity of Roadway Crashes in Tanzania

Boniphace Kutela, Ph.D. P.E
Assistant Research Scientist
Roadway Safety Program
Texas A&M Transportation Institute
701 N Post Oak Ln # 430, Houston, TX 77024
E-mail: b-kutela@tti.tamu.edu

Rafael John Mwekh'iga
Civil Engineer
Ibra Contractors Ltd
P.O.BOX 20881, Dar es Salaam, Tanzania
E-mail: rafaelmwekhiga@gmail.com

Arnold Mutegeki Kilaini
Civil Engineer
Road Asset Management & GIS Expert
P.O BOX 32405, Dar es Salaam, Tanzania
E-mail: kilaini1@gmail.com

Raynard Tom Magehema
Tutorial Assistant
Department of Civil Engineering
Ardhi University
P.O. Box 35176, Dar es salaam, Tanzania
E-mail: rmagehema@gmail.com

Geophrey Mbatta, Ph.D.
Highway and Traffic Engineering Lecturer
Department of Civil Engineering
Ardhi University
P.O.BOX 35176, Dar es Salaam, Tanzania
E-mail: mbattageo@gmail.com

Received: November 23, 2022 Accepted: December 15, 2022 Published: December 27, 2022
doi:10.5296/jss.v7i1.20488 URL: https://doi.org/10.5296/jss.v7i1.20488

Abstract

Traffic crashes are among the major cause of fatalities in most developing countries. A lack of a centralized historical crash database impedes the efforts to perform large-scale studies to understand the causes of crashes. However, the information for most major crashes are normally shared on social media. Therefore, this study used social media posts to extract information related to head-on and run-off roadway (ROR) crashes that occurred between 2010 and 2020. A total of 661 crashes were collected, which included 409 head-on and 252 ROR crashes. Geographical Information System (GIS) applications and multinomial logit models were developed to understand the spatial distribution of crashes and the resulting severity. It was found that the spatial distribution of crashes differs significantly by type and severity. Head-on collisions were predominant in two eastern regions, while ROR collisions appeared to spread on a large scale across the entire country. The pattern of fatal collision includes one southern region in addition to the two Eastern regions. Conversely, the pattern of non-injury crashes shows predominance along the coastal area. The Multinomial model results showed that speeding was two times more likely to result in fatalities for ROR crashes. Crashes in rural areas were about two times more likely to be fatal, given that they were head-on collisions. Further, bus-involved crashes were less likely to result in injuries. The implications of these study findings to the practitioners are discussed.

Keywords: Head-on and ROR crashes, social media, crash severities.

1. Background

Traffic crashes is the largest single cause of death for people aged between 5-29 years Worldwide. Most of the low- and middle-income countries located in Africa are severely affected by traffic crash fatalities. Statistics show that 93% of all fatalities occur in low and middle-income countries, which have only 60% of the total number of vehicles worldwide (WHO 2021). Among the cited reasons for the higher fatalities are poor visibility and visual guidance along the road (Ackaah, Apuseyine, and Afukaar 2020; Adanu et al. 2020; Gainewe and Masangu 2010). Road conditions, environmental conditions (Abdulrahman, Ibrahim, and Rogo Usman 2020; Adanu et al. 2020; Gainewe and Masangu 2010), and traffic-related factors (Gainewe and Masangu 2010; Muguro et al. 2020) have also been linked to fatalities. A study by Adanu et al. (2020) in Namibia that covered five years of crash data revealed that high occupancy passenger vehicles were associated with fatalities and severe injuries. Another South African-based study (Gainewe and Masangu 2010) revealed that speeding, unsafe overtaking, and slippery roads were the leading cause of fatal crashes.

Despite the higher fatalities, a relatively low number of traffic safety studies originate in developing countries, especially those in Africa, compared to other developing countries. For a country like Tanzania, even fewer nationwide studies are available (Boniface et al. 2016; Chiduo and Minja 2001). Most available studies are either corridor-based or focus on a specific region or location (Salum et al. 2019; Zimmerman et al. 2015). One of the reasons for the lack of large-scale studies is the difficulty in obtaining nationwide data from authorities. Until recently, Tanzania did not have a centralized database storing crash data from all regions (RAIS 2021). Before introducing the Road accident information system (RAIS), researchers who wanted to perform traffic safety studies would need to travel across all regions to collect crash data. Even with the presence of RAIS, the past decade's historical crash data are unavailable as the system started storing data in recent years.

Previous studies have indicated that head-on collisions and ROR crashes are common in most African countries (Abdulrahman et al. 2020; Adanu et al. 2020; Gainewe and Masangu 2010), including Tanzania. For instance, Francis et al. (2021) revealed that about 15% of crashes were head-on collisions. Another study by Salum et al. (2019) showed that 20 % of all vehicle-motorcycle crashes in Dar es Salaam, Tanzania, were head-on collisions. In South Africa, head-on collisions ranged between 8% and 11% between 2005 and 2009 (Gainewe and Masangu 2010). The same crash type was mentioned as one of the predominant in Kenya (Muguro et al. 2020) between 2015 and 2020. Likewise, the ROR crashes were reported in several studies in Tanzania. Kenya, South Africa, and Namibia, among other countries (Adanu et al. 2020; Muguro et al. 2020; Salum et al. 2019). The higher proportion of ROR and head-on crashes is associated with the roadway geometry across African countries. For instance, in Tanzania, most roads, especially in rural areas, are two-way, two-lane undivided roadways that increase the chance of head-on collisions. Further, the lack of wide shoulders and clearance zone increases the chance of run-off-roadway (ROR) crashes.

Although most developing countries lack a centralized database that stores crash data, crash data are normally shared through social media and stored in the cloud. Further, the rapid increase of the internet spread across developing countries, including Tanzania, has enabled millions of people to own mobile cell phones that can access the internet. For instance, mobile phone owners in Tanzania grew from 115,000 in 2000 to 23.14 million in 2020 (Ng'wanakilala 2019). Currently, about six out of ten Tanzanians do own a smartphone. Such an increase in smartphone owners has coupled with the number of people owning social media accounts. Most of these accounts are used to share various information, including severe vehicular crashes.

Several previous studies have used social media data (Ghandour et al. 2020; Ghandour, Hammoud, and Telesca 2019; Gu, Qian, and Chen 2016; Kutela, Magehema, et al. 2022; Zhang et al. 2018). For instance, Zhang et al. (2018) applied deep learning to detect traffic accidents from social media data in two metropolitan areas: Northern Virginia and New York City. Gu et al.

(2016) presented a methodology to extract publicly available tweets for free for traffic incident identification. Further, Ghandour et al. (2019) studied the spatial-clustering behavior and hazard vulnerability of car accidents in Lebanon between 2015 and 2018 using social media data. Additionally, Ghandour et al. (2020) proposed a real-time online platform to collect crash events from social media.

Irrespective of the origin and type of crashes, ordered and unordered models have predominantly been used to model the severities. A study by Saleem, Al-bdairi, and Hernandez (2020) used the latent class ordered probit models with two classes to investigate the effect of area type (urban vs. rural) on injury severity outcomes sustained by drivers in ROR crashes involving large trucks. Further, Rezapour and Ksaibati (2018) applied a multinomial regression model to identify the contributory factors to the risky violations and truck crashes. Hosseinpour, Yahaya, and Sadullah (2014) applied a random-effect generalized ordered probit model to analyze a head-on collision in Pakistan. In North Carolina, (Liu and Fan 2018) used a mixed logit model approach to evaluate contributing factors that significantly affect the severity of head-on crashes. Another study by Vachal et al., (2016) used multivariate models of three driver groups to identify the risk factors for rural severity of truck crashes. This study shows that several predictors are significantly associated with an increased likelihood of severe driver injury outcomes. Chen, Fan, and David (2019) in North Carolina applied a multinomial logit model for modeling pedestrian-vehicle crash severity. Similarly, Salum et al. (2019) applied a multinomial logit model for modeling motorcycle crashes in Dar es Salaam, Tanzania.

Geographical Information System (GIS) -based application has previously been used to explore the spread and the crash hotspots to understand the spatial coverage of crashes. For instance, a Salem et al. (2006) study explored the spatial distribution of crashes in the work zone areas using GIS applications. Similarly, Lee and Khattak (2019) and Emaasit et al. (2013) identified crash hotspots using GIS applications. However, applying GIS to analyze crashes' spatial distribution requires several crash information, including geolocation. In most cases, crashes in developing countries, especially Tanzania

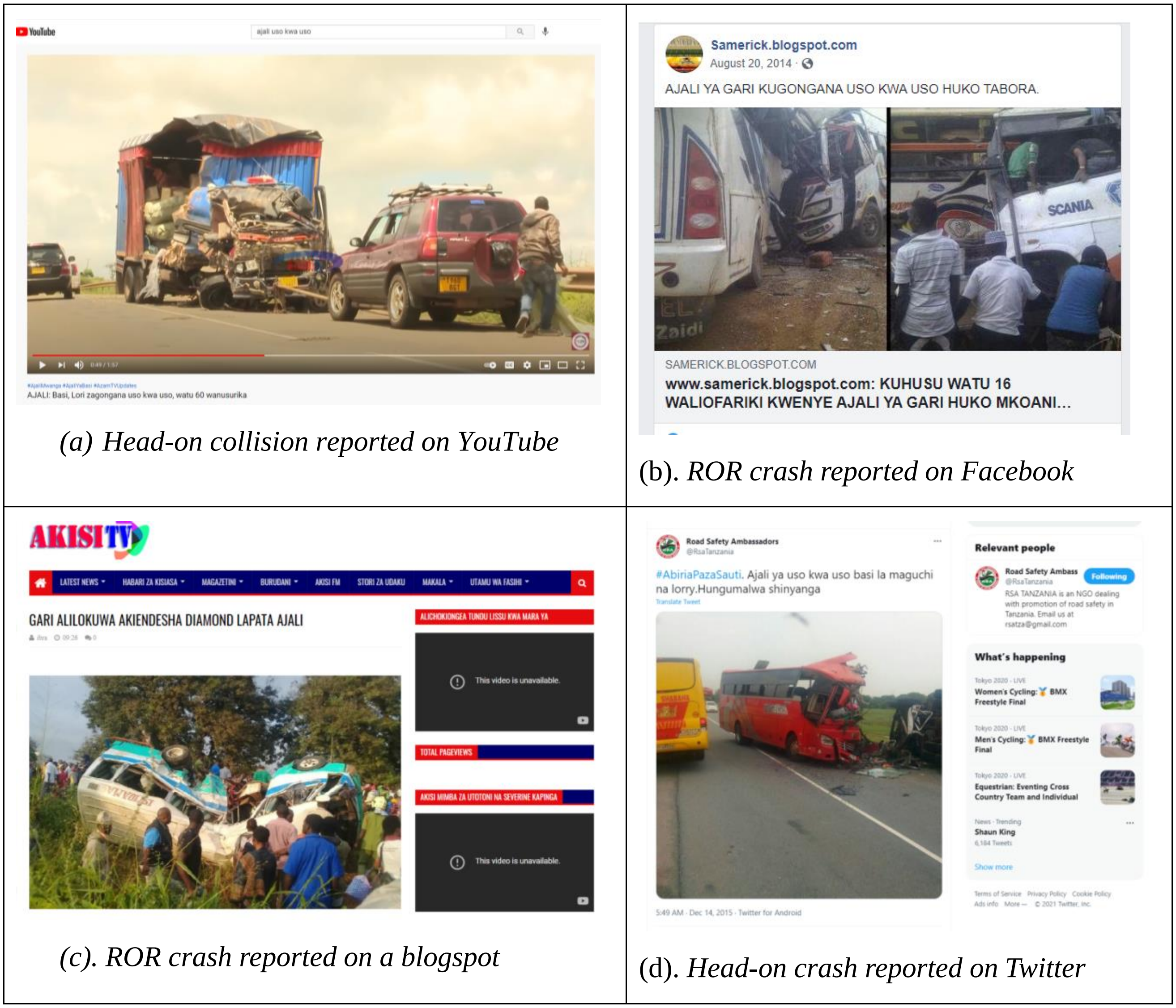
Therefore, the literature shows that traffic fatalities severely affect African countries, including Tanzania. Head-on collisions and ROR crashes are very common, especially in rural settings; however, few studies show the spatial distribution and associated severities, especially in Tanzania. The possible reason for few studies is a lack of a centralized data collection system with historical crash data. However, in most cases, crash information and detailed reports from witnesses or police officers are available online through several social media outlets. Therefore, this study leveraged crash information on social media outlets to analyze head-on and ROR crashes in Tanzania for the past decade. Specifically, the study focused on the spatial and severity analysis of the crashes reported through social media. Understanding the spatial coverage would help law enforcement officers to target specific locations to lower the number of crashes. On the other hand, severity analysis will enable an understanding of the key factors for injury severity. The rest of the paper is organized as follows; the next section presents the methodology and study data. The results and discussion are presented, followed by the conclusion and future studies.

2. Study Method

This section presents the methodological approach used in this study. It covers the data collection strategies, data extraction and processing, GIS mapping and heatmap development, and the statistical models.

2.1 Data Collection

As described earlier, this study used data that had been collected from various social media outlets. The research team manually searched crashes that were highlighted in social media for the past decade. The keywords “crash”, “accident”, and “ajali”, which translates to crash, were used on Twitter, Facebook, Instagram, and YouTube to capture the archived crash information. ***Figure* 1** presents screenshots of the typical social media posts showing traffic crashes.



(a) Head-on collision reported on YouTube

(b). *ROR crash reported on Facebook*

(c). ROR crash reported on a blogspot

(d). *Head-on crash reported on Twitter*

Figure 1. A screenshot of typical social media posts showing traffic crashes

Journal of Safety Studies
ISSN 2377-3219
2022, Vol. 7, No. 1

The associated crash narratives for each crash were extracted and stored in an excel sheet. Since one crash could have been reported by multiple sources, credible sources' narratives and rich information were selected for each crash. Google translator was used to translating the Swahili language to the English language. For instance, this Twitter post by Road Safety Ambassadors in Tanzania (@RsaTanzania 2021) "Ajali ya uso kwa uso basi la maguchi na lorry." meant "A face-to-face accident with a bus and a lorry".

2.2 Data Extraction and Processing

As the intention was to extract head-on and ROR crashes, the analyst reviewed all the narratives and extracted the respective crashes. Sometimes, the analyst had to watch the entire YouTube video to extract the necessary information. The information of interest included the number of people killed and injured, the number and type of vehicles involved, the date and time of the day, the crash location, possible causes of the crash, and the region and zone where the crash occurred. Upon completion of data extraction, a total of 661 crashes, which include 409 head-ons, and 252 ROR, were available for further analysis.

2.3 Mapping of Crashes

A python code that geolocates crash locations was prepared using the geopy library to map the crashes. This code used crash narratives as the inputs to geolocate the crashes. The code was linked to Open Street maps for easy geolocating of the crashes. The geolocated crashes were mapped on the GIS-based map overlayed on the road network for further analysis. **Figure 2** shows the distribution of geocoded crashes by type and road conditions. These crashes will be used to develop heatmaps that will explain the intensity of crashes by type and severity.

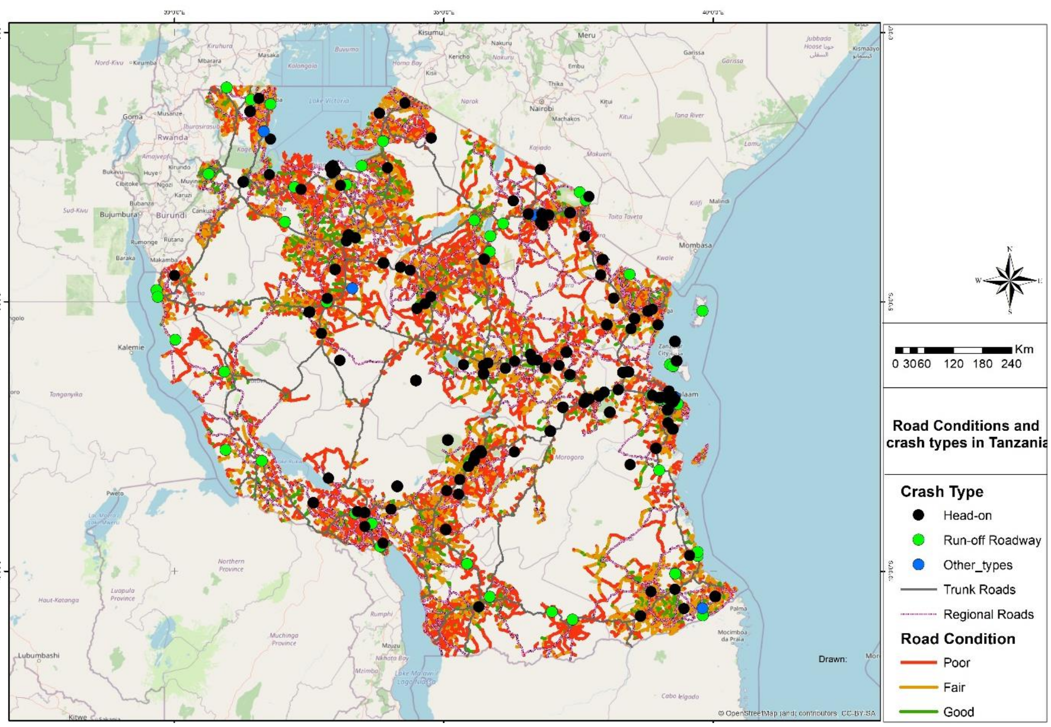


Figure 2. Spatial distribution of crashes by type and associated road condition

2.4 Statistical Modeling Methodology

The statistical model in this study intends to explain the influence of predictors on the response variable. The response variable for this case is the injury outcome, which is fatalities, injuries, and no injuries. By nature, injury outcomes are ordered; however, both ordered, and unordered models have been applied in the past (Chen et al. 2019; Liu and Fan 2018; Rezapour and Ksaibati 2018; Saleem et al. 2020; Salum et al. 2019). Thus, this study also applied the Multinomial Logit (MNL) model to understand the influence of various predictors on crash severity. Compared to the ordered models, the added advantage of this model is the ability to show the coefficients of all categories that are being compared to the base category.

The application of the MNL requires the selection of the base category (Chimba, Emaasit, and Kutela 2012; Kutela, Kidando, et al. 2022; Kutela, Kitali, et al. 2022; Liu et al. 2015; Salum et al. 2019). In this case, the category "no injuries" was used as the base category. Suppose that the injury outcome, Y has K total number of observed categories. The probability of each injury outcome category is given as:

$$Prob_{ik} = \frac{exp(\lambda_{ik})}{\sum_{k=1}^{K} exp(\lambda_{ik})} \tag{1}$$

where:

$$\lambda_{ik} = \beta_k + \sum_{k=1}^{K} \beta_{kj} X_{ki}$$

where β_k is the pre-crash-specific constant for behavior *k*, β_{kj} is a coefficient for variable *j* and behavior *k*, and X_i is the explanatory variable *k* for observation *i*. The coefficient estimations of the MNL model were performed using the maximum likelihood method implemented in an R statistical software (R Development Core Team 2021) using the

The significant factors influencing crash severity are discussed based on the relative risk ratio (RRR). The RRRs are risk factors that are calculated by exponentiation of the estimated coefficients. The outcome implicitly indicates the ratio between a particular category's predicted probabilities to the base category. In this case, the relative probability of the fatal or injury ($Y=i$) to the no injuries ($Y = 1$). When the estimated RRR of a variable is greater than 1 (RRR > 1), the risk ratio increases, while when it is less than 1 (RRR < 1), the risk ratio decreases relative to a base category (Çelik and Oktay 2014; Kutela and Teng 2018, 2020). A unit change in the explanatory factor leads to the RRR of the predicted category to change (increase or decrease) relative to the reference group, given the other variables in the model are held constant.

3. Descriptive Analysis

A descriptive analysis of crash data was carried out to explore the distribution of crashes across variables of interest. The analysis is presented by focusing on the trend and severity in relation to the collision type (head-on and run-off-roadway). Overall, a total of 661 crashes were collected for the period of eleven years (2010-2020).

3.1 Crash Trend

Figure 3 presents the trend of crashes from 2010 to 2020. Regardless of the year, there were more reported head-on crashes than ROR crashes, except for 2017 and 2012. The first three years of the last decade had fewer crashes reported on social media. The reported crashes peaked in 2014, declined thereafter, and increased in 2018. Social media reporters were increasing from 2013 onwards, and the most observation by the year 2014 was due to the increased number of social media reporters. The government made policies, law enforcement, and implementation to reduce the number of traffic crashes. Due to this, there were few crashes reported on social media in the three mid-decade years, and thereafter the rashes started to increase.

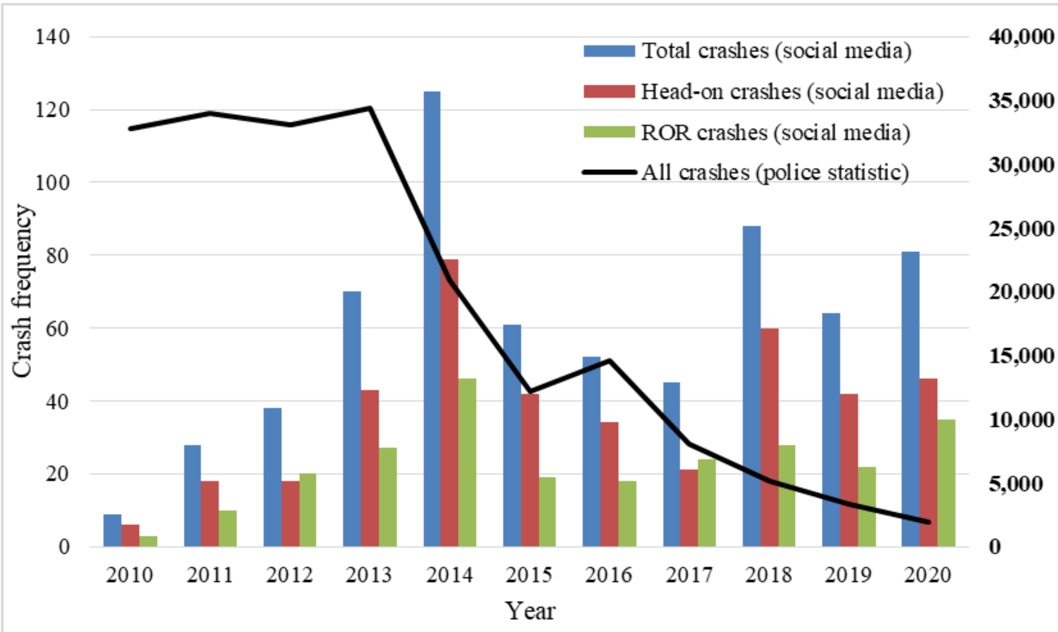


Figure 3. The trend of headlined crashes (Social media) and All Crashes (Police statistics)

Conversely, according to Figure 3, the trend is different for all crashes in Tanzania. The trend line shows that there is a progressive decrease in the trend from 2013 to 2020. The police reports show that the total number of crashes peaked from 2010 to 2013 and started to decrease to very few in 2020. This is different from the trend of headlined crashes, showing an increase in the reported crashes from 2018 to 2020.

3.2 Crash Severities

Furthermore, **Figure 4** shows the distribution of crash severity by collision type. It can be observed that more fatal crashes were reported compared to other injury or PDO crashes. Although this might sound counterintuitive, it should be noted that the analyzed data were collected from the headlined crashes on social media. Social media influencers are less likely to report PDO crashes as they will not draw followers' attention. For social media crashes, head-on collisions have caused more fatal crashes than injuries and PDOs. This might be due to the road conditions in Tanzania, most of which are two-way roadways with no median and a high posted speed limit. On the other hand, ROR crashes had a smaller number of observations compared to head-on. The presence of shoulder rumble strips on most of the rural highways in Tanzania can explain the low number of ROR crashes.

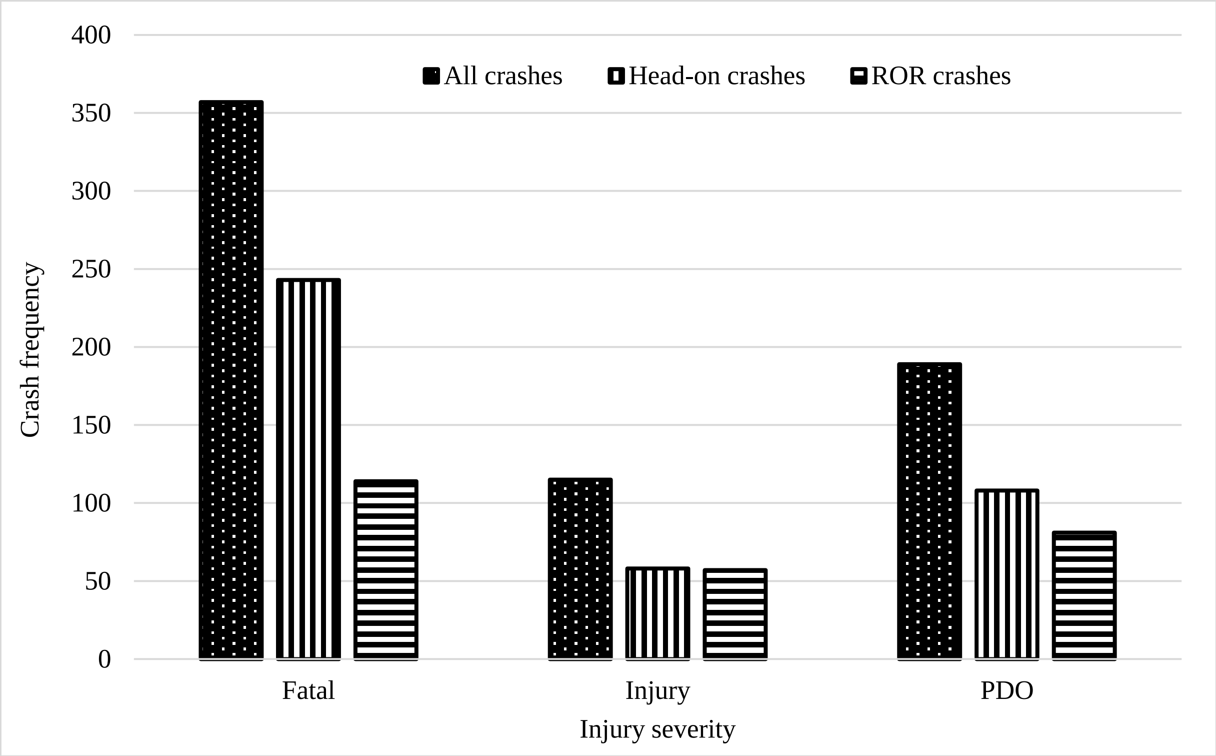


Figure 4. Distribution of crash severity by collision type

3.3 Descriptive Summary of the Variables

Table 1 presents crash types and severity distribution per variables of interest. Speeding is the leading cause of fatal crashes in both head-on and ROR crashes. It was also shown that more crashes occur on weekdays than on weekends for both head-on and ROR crashes. Crashes involving buses account for at least 58% of fatal crashes for head-on and ROR crashes. This variable indicates that buses are involved in as many crashes compared to other vehicles. On the other hand, truck-involved crashes comprise all crashes that a truck took part but did not involve buses. Also, car-involved crashes do not include any crash involving either a bus or a truck. The percentage of crashes with fire outbreaks is less than 10% for both fatal injuries and PDOs on head-on collisions. Similarly, for run-off-roadway crashes, the percentage of crashes that resulted in a fire was 16% for PDOs. The location of the crash was also a variable of interest. Most crashes occurred in rural areas than in urban areas for fatal injury and PDOs on both head-on and run-off-roadway crashes.

Table 1. Descriptive Analysis of the Variables

	Head-on crashes					
	Fatal		**Injury**		**PDO**	
	Count	Percent	Count	Percent	Count	Percent
Contributing factor						

Speeding	35	14.4%	3	5.2%	19	17.6%
Others	208	85.6%	55	94.8%	89	82.4%
Day of week						
Weekends	64	26.3%	8	13.8%	32	29.6%
Weekdays	179	73.7%	50	86.2%	76	70.4%
Vehicle involved						
Buses	141	58.0%	27	46.6%	75	69.4%
Cars	20	8.2%	19	32.8%	14	13.0%
Trucks	82	33.7%	12	20.7%	19	17.6%
Vehicle fire						
Yes	14	5.8%	2	3.4%	7	6.5%
No	229	94.2%	56	96.6%	99	91.7%
Crash location						
Rural	197	81.1%	39	67.2%	70	64.8%
Urban	46	18.9%	19	32.8%	38	35.2%
			Run-off-Roadway Crashes			
Contributing factor						
Speeding	24	21.1%	8	14.0%	8	9.9%
Others	90	78.9%	49	86.0%	73	90.1%
Day of week						
Weekends	27	23.7%	20	35.1%	24	29.6%
Weekdays	87	76.3%	37	64.9%	57	70.4%
Vehicle involved						
Buses	67	58.8%	20	35.1%	38	46.9%
Cars	28	24.6%	31	54.4%	18	22.2%
Trucks	19	16.7%	6	10.5%	25	30.9%
Vehicle fire						
Yes	5	4.4%	3	5.3%	13	16.0%
No	110	96.5%	56	98.2%	65	80.2%
Crash location						
Rural	99	86.8%	41	71.9%	68	84.0%
Urban	15	13.2%	16	28.1%	13	16.0%

4. Results and Discussion

This section presents the results and discussion. The section is divided into three sections: major sections, spatial analysis and discussion by crash type, spatial analysis and discussion by crash severity, and statistical model results and discussion.

4.1 Spatial Analysis and Discussion by Crash Type

The spatial analysis and discussion by crash type section cover head-on and ROR crashes. **Figure 5** and **Figure 6** present heatmaps for head-on and ROR crashes. According to **Figure 5**, the Dar es Salaam and Morogoro regions have a high number of head-on collisions. These two regions have had at least 59 crashes each for the past decade. It should be noted that these crashes are only the ones that were published by social media, meaning they caught people's attention. Two reasons can explain the higher number of reported head-on collisions in the two regions. First, for the Dar es salaam region, the higher number of reported crashes may be attributed to many vehicles within the city and the presence of many people with social media accounts. Thus even for not-so-serious crashes, people will publish them on social media. On the other hand, the Morogoro region has neither of the two factors mentioned for the Dar es salaam region. However, the possible reason for such a high number of reported crashes could be those involving buses and trucks traveling upcountry. This region has an extended network of trunk roads and two-way two-lane highways. The chance of head-on collision on these types of roadways is very high. The remaining regions have a relatively small number of crashes.

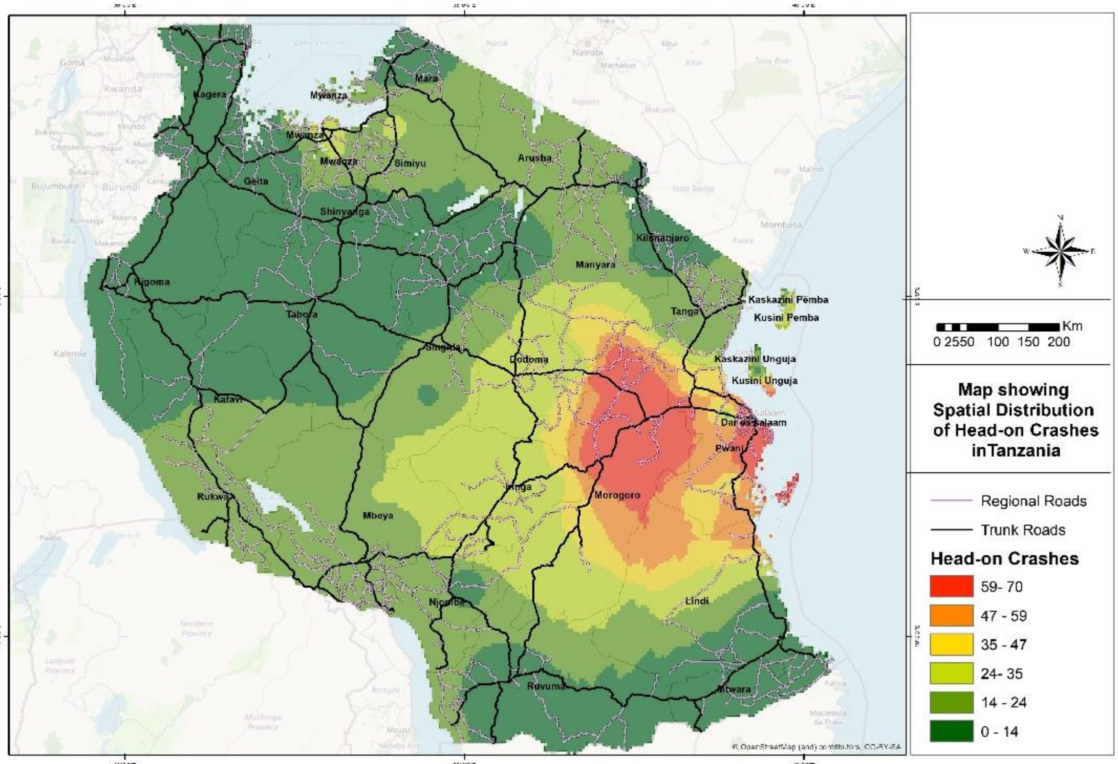


Figure 5. A heatmap of head-on crashes

The ROR crashes (**Figure 6**), on the other hand, have extended to several other regions other than Dar es Salaam and Morogoro. It can be observed that although there observed a high number of ROR crashes in dar es salaam, the spread is relatively small. This is because most of the roadways in Dar es salaam have paved shoulders, although the two-lane roadways are still dominant. On the other hand, the Morogoro region still shows extensive coverage of ROR crashes. This is because fewer roads have paved shoulders and have a large network of this type of roadway. The intensity of ROR crashes extends to other regions such as Tabora, Arusha, and

Kagera. These regions have a relatively low number of vehicles, but the road conditions are poor.

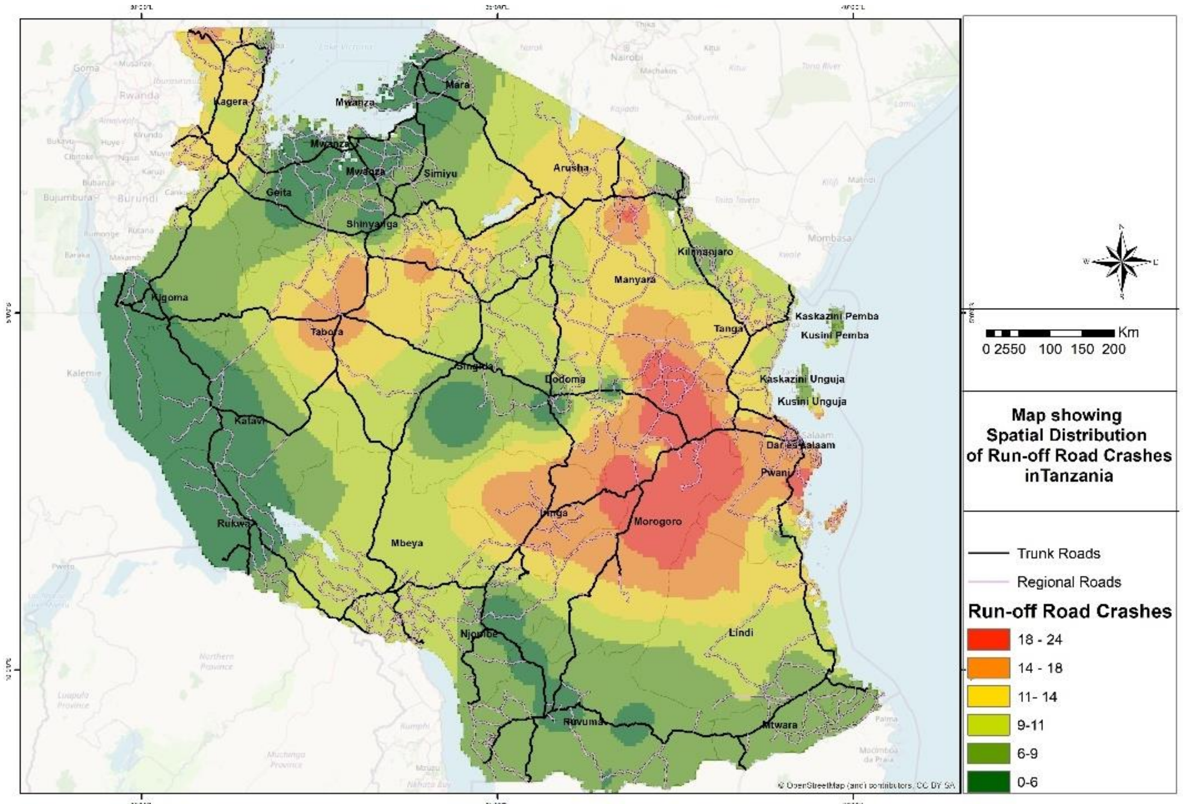


Figure 6. A heatmap of ROR crashes

4.2 Spatial Analysis and Discussion by Severity

The spatial analysis and discussion by crash severity section cover fatal, injury, and PDO crashes. **Figure 7**, **Figure 8**, and **Figure 9** present heatmaps for fatal, injury, and PDO crashes. According to **Figure 7**, social media reported that fatal crashes are dominant in the Morogoro and Lindi regions. As described earlier, the Morogoro region had the highest proportion of fatal head-on collisions. Therefore it is likely that these head-on collisions resulted in fatal crashes. Other regions, such as Arusha, mara, Iringa, and Dar es Salaam, had relatively fewer fatal crashes reported on social media compared to Morogoro and Lindi. The southern region has been shown to have a high number of crashes. In the past, several cement-carrying trucks were involved in several crashes. Since they were under the same owner, there was a myth that drivers were boycotting, so why do they cause crashes. However, the situation was then resolved. The injury crashes in **Figure 8** show a different pattern from fatal crashes. It can be observed that Dar es salaam is the region with a high number of injury crashes. The high volume of vehicles which caused slower speeds in the city, maybe the main reason for this observation. The pattern of injury crashes in the Morogoro region is more aligned to the regional roads than trunk roads. The regional roads have relatively lower speed limits compared to trunk roads which increases the chance of injury crashes. Furthermore, the northwest of the country (Kagera region) has shown a high injury crash density pattern.

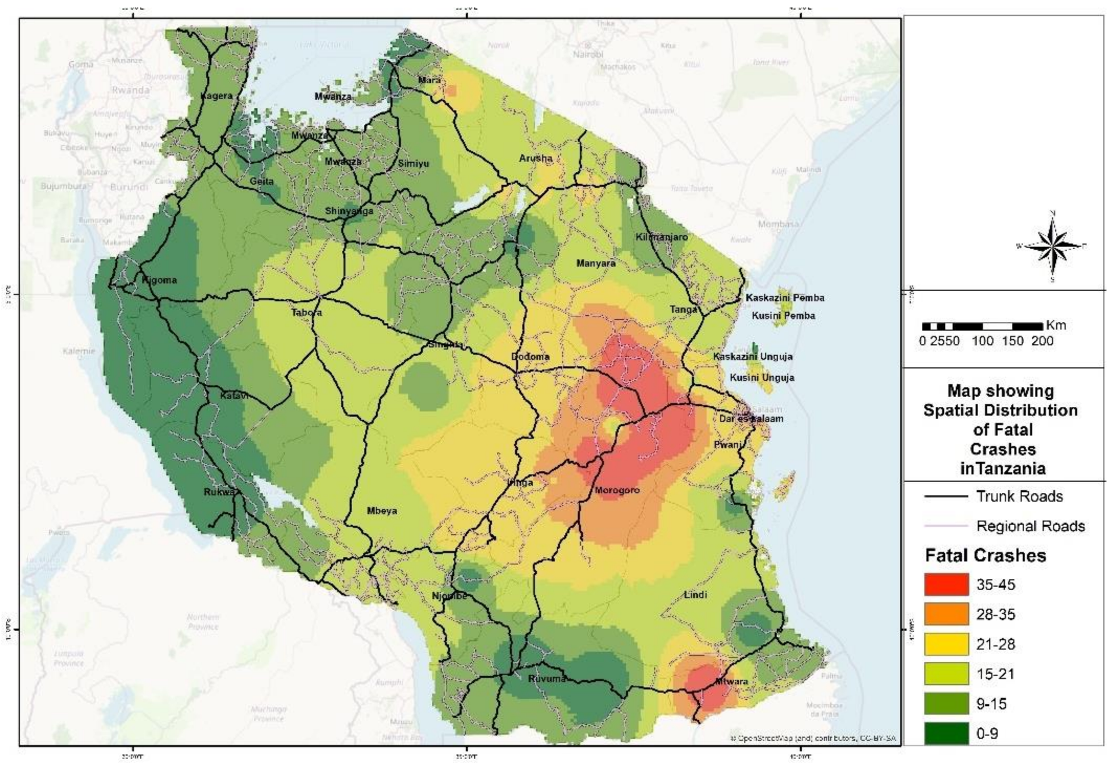


Figure 7. A heatmap of fatal crashes

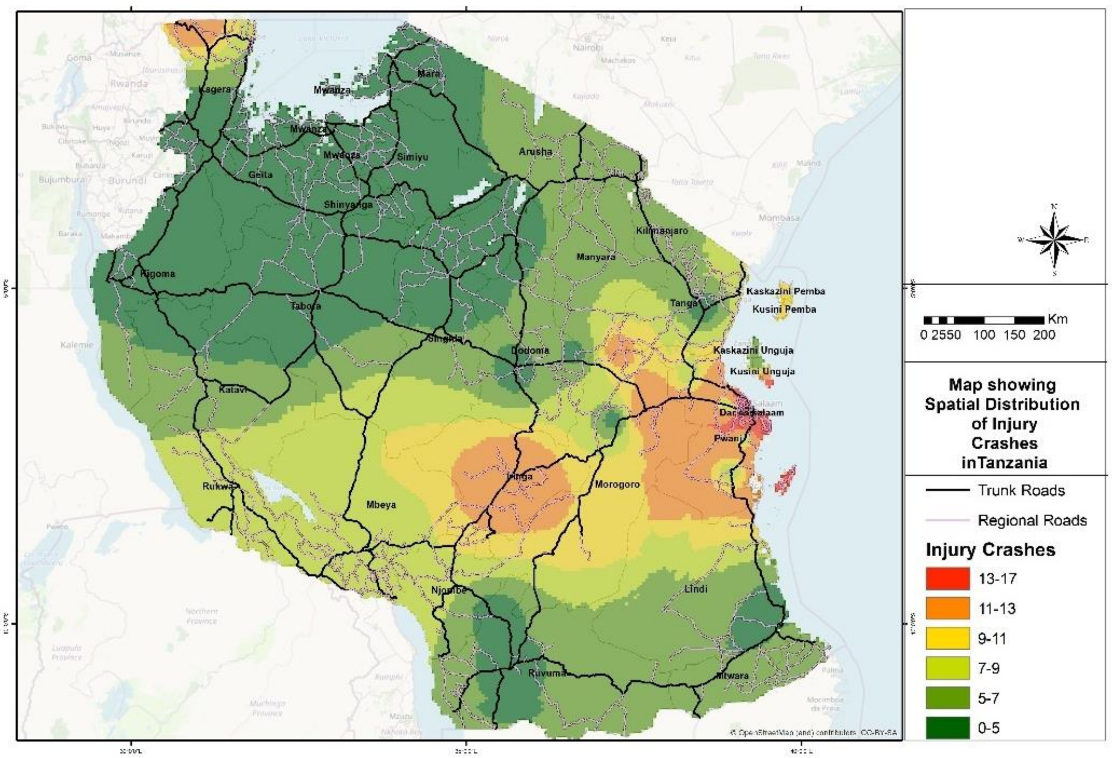


Figure 8. A heatmap of injury crashes

Lastly, the PDO crashes presented in **Figure 9** portray a different pattern compared to the fatal and injury crashes. Reported PDO crashes are observed to be more in the coastal regions, including Dar es Salaam, Tanga, and Zanzibar. For Dar es salaam, the possible reason for dense reported PDO crashes is the higher number of vehicles that cause congestion. Most of the time, crashes that occur in congested areas are likely to be PDOs. A similar scenario can also explain the crash pattern in Zanzibar.

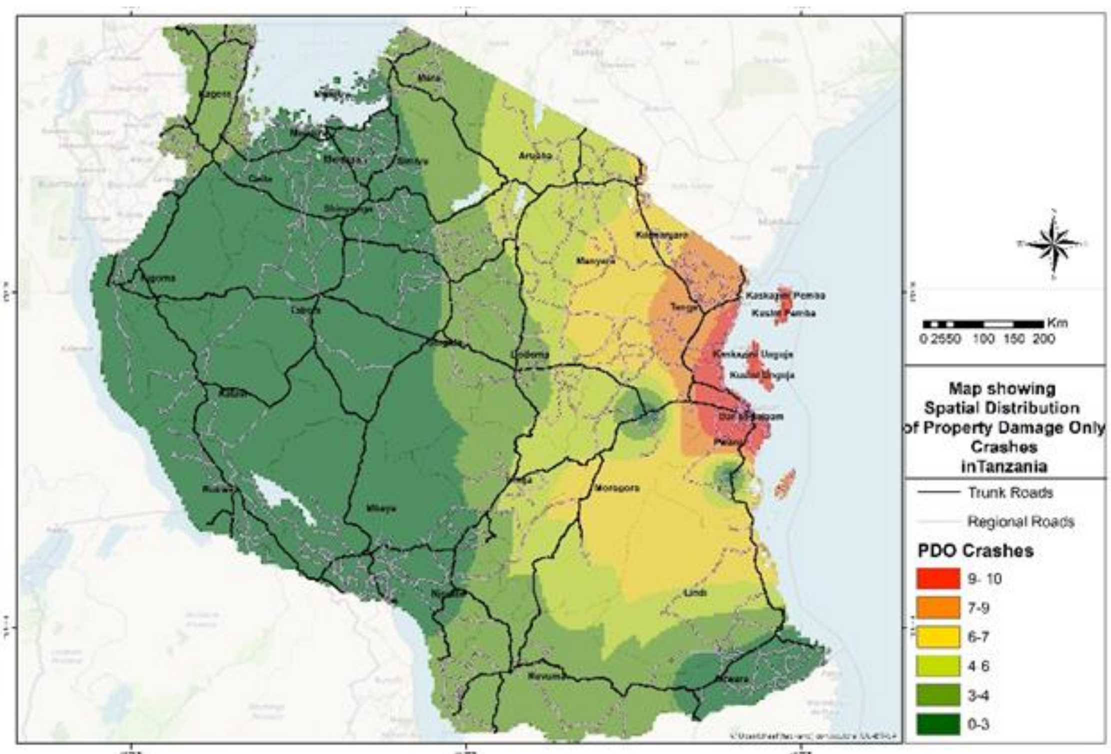


Figure 9. A heatmap of PDO crashes

4.3 Statistical Model Results and Discussion.

Before developing the regression model, correlation analysis was performed to understand the correlation across the independent variables. Highly correlated variables tends to result into multicollinearity (Kutela 2022; Tu et al. 2005). Therefore, a correlation coefficient greater than 0.5 was deemed high. According to **Figure 10**, the variables bus involved and truck involved were highly correlated with a coefficient of 0.64, and the variables bus involved and car involved were highly correlated with a correlation coefficient of 0.55. Other variables like speeding, weekend, and vehicle fire have relatively low correlation coefficients. Therefore, all variables were considered in the model.

4.3.1 Multinomial Logistic Regression model results

The results of the multinomial logistic regression model are displayed in **Table 2**. The discussion and interpretation of the model were based on the 90% confidence interval, which shows the significance of the predictors in relation to either injury or fatal crashes and the base outcome (PDOs). The six predictors that have been discussed in this model are the contribution factor, vehicles involved, fire outbreak, crash location, day of the week, and the type of collision. contribution

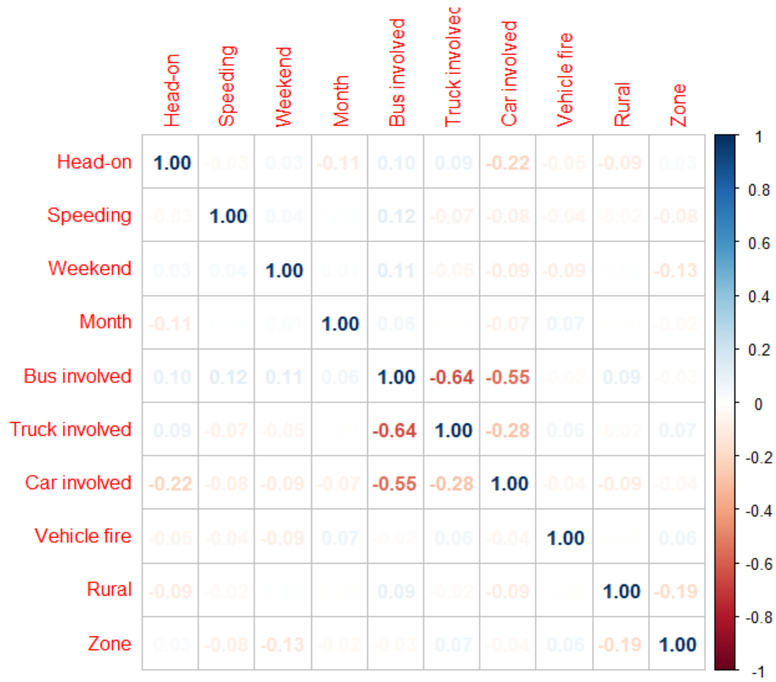


Figure 10. Correlation of the variables for the MNL model

4.3.1.1 Collision type

Results in **Table 2** show that head-on collision are less likely to result in injuries but more likely to be PDO. The odds of fatal crashes occurring in head-on collisions increase by about 62% in relation to PDOs. The findings are in line with the previous studies indicated that more than two-thirds of the fatal crashes involved head-on than other crash types. Furthermore, the study by Musa, Hassan, and Mashros (2020) showed that the likelihood of more serious accident severity due to poor horizontal alignment was correspondingly about 0.4 times less compared to the absence of such factors. However, this study shows that injury crashes are not statistically significant at a 95% confidence level for run-off-roadway crashes.

4.3.1.2 Speeding

Speeding is one of the factors of interest in crash severity. The results in **Table 2** show that speeding was less likely to result in fatal crashes than PDOs when the collision was head-on and more likely to cause fatal crashes than PDOs when the collision was run-off-roadway. This is contrary to a previous study by Malyshkina and Mannering (2014), which showed that the speed limit did not have a statistically significant effect on the severity of accidents on interstate highways. Conversely, the study by Farmer (2019) showed that a five mph increase in the maximum state speed limit was associated with an 8.5% increase in fatality rates on interstates/freeways

Table 2. Multinomial logistic regression model results

	All crashes						Head-on crashes						Run-off-roadway crashes					
	Injury			**Fatal**			**Injury**			**Fatal**			**Injury**			**Fatal**		
Predictor	Coeff	RRR	P-value	Coeff	RRR	P-value	Coeff	RRR	P-value	Coeff	RRR	P-value	Coeff	RRR	P-value	Coeff	RRR	P-value
Contributing factor																		
Other factors																		
Speeding	-0.38	0.68	0.325	0.21	1.23	0.416	-1.17	0.31	0.076	-0.06	0.94	0.860	0.62	1.86	0.268	0.86	2.36	0.058
Vehicle involved																		
Cars vs non-bus/truck																		
Bus vs any other	-1.31	0.27	<0.001	0.001	1.00	0.996	-1.48	0.23	0.001	0.15	1.17	0.695	-1.01	0.37	0.018	0.14	1.15	0.718
Trucks vs non-bus	-1.26	0.28	0.001	0.32	1.38	0.292	-0.87	0.42	0.103	1.04	2.82	0.018	-1.97	0.14	0.000	-0.73	0.48	0.101
Fire outbreak																		
No																		
Yes	-2.81	0.06	0.007	-1.01	0.36	0.002	-11.09	0.00	0.919	-0.51	0.60	0.279	-2.69	0.07	0.013	-1.88	0.15	0.002
Crash location																		
Urban																		
Rural	-0.09	0.92	0.756	0.674	1.96	0.003	0.19	1.21	0.607	0.85	2.34	0.001	-0.83	0.43	0.082	-0.03	0.97	0.941

Day of the week																		
Weekend																		
Weekday	0.361	1.44	0.207	0.145	1.16	0.487	1.18	3.24	0.011	0.18	1.20	0.502	-0.28	0.76	0.492	0.21	1.24	0.542
Collision type																		
Run-off-roadway																		
Head-on	-0.1	0.90	0.697	0.483	1.62	0.013												
Intercept	0.45	1.57	0.229	-0.33	0.72	0.321	-0.44	0.65	0.412	-0.26	0.77	0.541	1.43	4.19	0.008	0.36	1.43	0.493
Observations	661						409						252					
AIC	1257						734						510					
BIC	1329						791						560					

and a 2.8% increase on other roads. In total, during the 25-year study period, there were an estimated 36,760 more traffic fatalities than would have been expected if maximum speed limits had not increased, 13,638 on interstates or freeways and 23,122 on other roads. However, from the findings of this study, speeding is not statically significant at 95% in fatal crashes for run-off-roadway crashes.

4.3.1.3 Type of vehicle involved

From the model results in **Table 2**, the chance of the buses resulting in injury crashes appeared to decrease relatively compared to PDOs. The odds of the buses to result in fatal crashes appeared to increase compared to PDOs for both collisions. Also, compared to base outcome (PDOs), the bus involved was less likely to result in injury crashes and more likely to result in fatal crashes. The findings are in hand with a study by Linfeng Won (2005), which showed that older drivers are less likely to experience possible injuries in a large-size vehicle (truck or bus) compared to passenger cars. However, the bus-involved crashes were not statically significant at a 95% confidence interval for injury crashes.

Like buses, trucks were less likely to result in injury crashes than PDOs and more likely to result in fatal crashes than PDOs for all types of collisions. For head-on collisions, the odds of the trucks resulting in injury crashes decreased by 58%, while the odds of trucks causing fatal crashes increased by 4% compared to the base outcome. In run-off-roadway crashes, the odds of injuries caused by trucks decrease by 86% compared to PDOs. Other previous study showed that truck has resulted in more fatal crashes than PDOs due to the traffic increase (Zwerling et al. 2005). Generally, trucks involved had a high effect on fatal crashes compared to PDOs for head-on collisions

4.3.1.4 Fire outbreak

Fire outbreak is among the factors of interest in the discussion. From **Table 2**, a fire outbreak was less likely to result in injury than PDOs. The chance of a fire outbreak in a fatal crash decreased by 40% compared to PDOs in head-on collisions. This implies that when a fire outbreak occurs, there is a low chance of the crash being fatal or injured and a high chance of being PDO. However, fire outbreak was not statically significant at a 95% confidence interval on injury and fatal crashes. Similary, previous studies showed that crashes involving fire outbreaks are less likely to result in fatal crashes than property damage only (NHTSA 1990). Furthermore, fatalities were not affected by fire outbreaks. Also, fires in fatal passenger car crashes have increased, primarily due to an aging vehicle fleet. Fire risk increases with vehicle age, and the contributions of fire and impact forces in causing death and injury cannot be separated in crash data files (NHTSA 1990).

4.3.1.5 Crash location

The results in **Table 2** show that in rural areas, fatal crashes were more likely to occur than PDO crashes for head-on collisions. Compared to PDOs, the chance of occurrence of injury crashes in head-on collisions increased by 21%, while the chance of occurrence of fatal crashes in rural areas was more than twice the occurrence of injury crashes. Both injury and fatal crashes have less chance of occurrence in rural areas than the PDOs in Run-off-roadway collisions. The same results have been reported from previous studies (Vachal et al. 2016), showing that rural areas have more fatal crashes than PDOs. Rural areas are not statically significant at a 95% confidence interval on fatal crashes for head-on collisions. On the other hand, these results do not agree with Aghajani et al. (2017), which concluded that the roads northwest of Ilam Province-Iran, predominantly rural areas, have less traffic higher number of fatalities.

4.3.1.6 Day of the week

The day of the week was among the variables of interest. Overall, weekday crashes are observed to have a higher likelihood of resulting in either fatal or injury crashes than PDOs. However, only head-on collisions have shown statistically significant results and a 95% confidence level. The chance of injury and fatal crashes increase by 44% and 16%, respectively, for all crashes. Conversely, for run-off-roadway crashes, the chance of injury crashes decreased by 24%, while fatal crashes increased by the same magnitude. Furthermore, the odds of injury crashes are more than three times for head-on collisions during weekdays. The study by Li, Song, and Fan (2021) have explained the likelihood of the increased odds of fatal head-on crashes during weekends. Also, another study by Shrestha and Shrestha (2017) showed that crashes on weekdays were three times more likely to be injury crashes than on weekends.

5. Conclusion and Future Studies

This study mainly aimed at presenting the spatial and severity analysis of head-on and run-off-roadway crashes in Tanzania based on social media data. Initially, a data description is presented, followed by spatial analysis and statistical modeling using the multinomial logistic regression model.

The descriptive analysis shows an increasing trend up to 2014 due to the increasing number of social media reporters. Conversely, the police report of all crash data shows a progressive decrease in the trend, particularly from 2013. Also, social media presents more fatal crashes than PDO crashes since fatal crashes are likely to trend more on social media than PDOs. Further, there have been more head-on crashes than ROR crashes since the nature of Tanzania roads is mostly two-way roadways, without medians, causing vehicles to collide head-to-head. Fatal crashes appear to have a higher frequency than other severities. Therefore, road users must be educated on transportation safety to reduce these effects.

On the other hand, the spatial distribution of the crashes in Tanzania revealed that the coastal zone requires special attention as it was mostly affected by headlined crashes. From social media achieved data, both head-on and ROR crashes have occurred mostly in the coastal regions. Despite the ongoing improvements in road conditions, policies and rules, particularly in road safety, need to be made and implemented to mitigate the problem.

Statistical modeling was performed for the associated variables affecting the severity of crashes. Multinomial logistic regression results show that speeding was more likely to cause injuries and fatalities, while the chances are high that buses will result in injury crashes compared to PDOs. Fire outbreak was less likely to cause injuries than PDOs. Rural areas had higher chances of fatal crashes occurring than PDOs. Finally, there is a high likelihood of fatal and injury crashes appearing on weekdays. Other crash causes have their impacts on injury and fatal severities.

One of the measures placed by authorities to reduce crashes in Tanzania includes speed torches and speed governors in buses. More research can be done to find out the effectiveness of these measures and review the general public’s perception of these measures using social media

Although the approach used in this study is of promise, more effort is needed to understand the nature, causes, and injury outcomes of all crashes. This approach was able to analyze only crashes that were reported on social media. It is known that most of the social media influencers live in the cities. Therefore, most of the events that happen in rural areas go unnoticed. A study that would use the data from the recently established database for storing crash data is recommended. Another limitation is the study data. Since Social media influences living in cities, crashes that occur in the villages are less likely to be reported.

References

@RsaTanzania. (2021). *AbiriaPazaSauti. Ajali Ya Uso Kwa Uso Basi La Maguchi Na Lorry.* Retrieved July 31, 2021, from https://twitter.com/RsaTanzania/status/676368507060297728

Abdulrahman, S., Ibrahim, K., & Usman, K. R. (2020). A Comparison of the Nigerian Road Traffic Crash Report Form with Other Selected Countries. *Bima Journal of Science and Technology*.

Ackaah, W., Apuseyine, B. A., & Afukaar, F. K. (2020). Road Traffic Crashes at Night-Time: Characteristics and Risk Factors. *International Journal of Injury Control and Safety Promotion, 27*(3), 392-399. https://doi.org/10.1080/17457300.2020.1785508

Adanu, E. K., Riehle, I., Odero, K., & Jones, S. (2020). An Analysis of Risk Factors Associated with Road Crash Severities in Namibia. *International Journal of Injury Control and Safety Promotion, 27*(3), 293-299. https://doi.org/10.1080/17457300.2020.1774617

Aghajani, M. A., Dezfoulian, R. S., Arjroody, A. R., & Rezaei, M. (2017). Applying GIS to Identify the Spatial and Temporal Patterns of Road Accidents Using Spatial Statistics (Case Study: Ilam Province, Iran). *Transportation Research Procedia*, *25,* 2126-2138. https://doi.org/10.1016/j.trpro.2017.05.409

Boniface, R., Museru, L., Kiloloma, O., & Munthali, V. (2016). Factors Associated with Road Traffic Injuries in Tanzania. *Pan African Medical Journal, 23*(1). https://doi.org/10.11604/pamj.2016.23.46.7487

Çelik, A. K., & Oktay, E. (2014). A Multinomial Logit Analysis of Risk Factors Influencing Road Traffic Injury Severities in the Erzurum and Kars Provinces of Turkey. *Accident Analysis and Prevention, 72,* 66-77. https://doi.org/10.1016/j.aap.2014.06.010

Chen, Z., Fan, W. D., & Wei, D. (2019). A Multinomial Logit Model of Pedestrian-Vehicle Crash Severity in North Carolina. *International Journal of Transportation Science and*

Technology, 8(1), 43-52. https://doi.org/10.1016/j.ijtst.2018.10.001

Chiduo, C. W., & Minja, P. (2001). Road Safety in Tanzania: What Are the Problems? In *First Road Transportation Technology Transfer Conference in Africa*.

Chimba, D., Emaasit, D., & Kutela, B. (2012). Likelihood Parameterization of Bicycle Crash Injury Severities. *Journal of Transportation Technologies, 2*(3), 213-219. https://doi.org/10.4236/jtts.2012.23023

Emaasit, D., Chimba, D., Cherry, C. R., Kutela, B., & Wilson, J. (2013). *A Methodology to Identify Factors Associated with Pedestrian High Crash Clusters Using GIS Based Local Spatial Autocorrelation.*

Farmer, C. M. (2019). *The Effects of Higher Speed Limits on Traffic Fatalities in the United States, 1993-2017.*

Francis, F., Moshiro, C., Yngve, B. H., & Hasselberg, M. (2021). Investigation of Road Infrastructure and Traffic Density Attributes at High-Risk Locations for Motorcycle-Related Injuries Using Multiple Correspondence and Cluster Analysis in Urban Tanzania. *International Journal of Injury Control and Safety Promotion*. https://doi.org/10.1080/17457300.2021.1930060

Gainewe, M., & Masangu, N. (2010). Factors Leading to Fatal Crashes and Fatalities on the South African Roads: 2005-2009. In *Proceedings of the 29th Southern African Transport Conference (SATC 2010)* (pp. 978-979).

Ghandour, A. J., Hammoud, H., Dimassi, M., Krayem, H., Haydar, J., & Issa, A. (2020). Allometric Scaling of Road Accidents Using Social Media Crowd-Sourced Data. *Physica A: Statistical Mechanics and Its Applications, 545,* 123534. https://doi.org/10.1016/j.physa.2019.123534

Ghandour, A. J., Hammoud, H., & Telesca, L. (2019). Transportation Hazard Spatial Analysis Using Crowd-Sourced Social Network Data. *Physica A: Statistical Mechanics and Its Applications, 520,* 309-316. https://doi.org/10.1016/j.physa.2019.01.025

Gu, Y. M., Qian, Z., & Chen, F. (2016). From Twitter to Detector: Real-Time Traffic Incident Detection Using Social Media Data. *Transportation Research Part C: Emerging Technologies, 67,* 321-342. https://doi.org/10.1016/j.trc.2016.02.011

Hosseinpour, M., Yahaya, A. S., & Sadullah, A. F. (2014). Exploring the Effects of Roadway Characteristics on the Frequency and Severity of Head-on Crashes: Case Studies from Malaysian Federal Roads. *Accident Analysis & Prevention, 62,* 209-222. https://doi.org/10.1016/j.aap.2013.10.001

Kutela, B. (2022). "The Role of Crosswalk-Related Features on Drivers' Spatial Yielding Compliance at Signalized Midblock Crosswalks." *Journal of Traffic and Transportation Engineering (English Edition)*. https://doi.org/10.1016/j.jtte.2021.11.001

Kutela, B., Kidando, E., Kitali, A. E., Mwende, S., Langa, N., & Novat, N. (2022). Exploring

Pre-Crash Gate Violations Behaviors of Drivers at Highway-Rail Grade Crossings Using a Mixed Multinomial Logit Model. *International Journal of Injury Control and Safety Promotion*. https://doi.org/10.1080/17457300.2021.1990348

Kutela, B., Kitali, A. E., Kidando, E., Mbuya, C., & Langa, N. (2022). Exploring the Need to Model Severity of Single- and Multi-Occupant Vehicles Crashes Separately: A Case of Crashes at Highway-Rail Grade Crossings. *International Journal of Transportation Science and Technology*. https://doi.org/10.1016/j.ijtst.2022.11.002

Kutela, B., Magehema, R. T., Langa, N., Steven, F., & Mwekh'iga, R. J. (2022). A Comparative Analysis of Followers' Engagements on Bilingual Tweets Using Regression-Text Mining Approach. A Case of Tanzanian-Based Airlines. *International Journal of Information Management Data Insights, 2*(2), 100123. https://doi.org/10.1016/j.jjimei.2022.100123

Kutela, B., & Teng, H. L. (2018). Parameterizing the Yielding Compliance of Motorists at Signalized Midblock Crosswalks Using Mixed Effects Logistic Regression. In *Transportation Research Board 97th Annual Meeting*. Washington DC, United States.

Kutela, B., & Teng, H. L. (2020). *Assessment of Methodological Alternatives for Modeling the Spatiotemporal Crossing Compliance of Pedestrians at Signalized Midblock Crosswalks.* https://doi.org/10.1061/JTEPBS.0000300

Lee, M., & Khattak, A. J. (2019). Case Study of Crash Severity Spatial Pattern Identification in Hot Spot Analysis: Transportation Research Record. *Journal of the Transportation Research Board*. https://doi.org/10.1177/0361198119845367

Li, Y., Song, L., & Fan, W. D. (2021). Day-of-the-Week Variations and Temporal Instability of Factors Influencing Pedestrian Injury Severity in Pedestrian-Vehicle Crashes: A Random Parameters Logit Approach with Heterogeneity in Means and Variances. *Analytic Methods in Accident Research, 29*(December 2020), 2020-2021. https://doi.org/10.1016/j.amar.2020.100152

Won, L. F., & Fan, W. D. (2005). *Modeling Single-Vehicle Run-off-Road Crash Severity in Rural Areas: Accounting for Unobserved Heterogeneity and Age Difference.*

Liu, J., Khattak, A. J., Richards, S. H., & Nambisan, S. (2015). What Are the Differences in Driver Injury Outcomes at Highway-Rail Grade Crossings? Untangling the Role of Pre-Crash Behaviors. *Accident Analysis and Prevention, 85*, 157-169. https://doi.org/10.1016/j.aap.2015.09.004

Liu, P. F., & Fan, W. D. (2018). Modeling Head-on Crash Severity on NCDOT Freeways: A Mixed Logit Model Approach. *Canadian Journal of Civil Engineering, 46*(6), 322-328. https://doi.org/10.1139/cjce-2018-0262

Malyshkina, N. V., &Mannering, F. L. (2014). Analysis of the Effect of Speed Limit Increases on Accident-Injury Severities. *ArXiv* (July).

Muguro, J. K., Sasaki, M., Matsushita, K., & Njeri, W. (2020). Trend Analysis and Fatality Causes in Kenyan Roads: A Review of Road Traffic Accident Data between 2015 and

2020. *Cogent Engineering, 7*(1). https://doi.org/10.1080/23311916.2020.1797981

Musa, M. F., Hassan, S. A., & Mashros, N. (2020). The Impact of Roadway Conditions towards Accident Severity on Federal Roads in Malaysia. *PLoS ONE, 15*, 1-12. https://doi.org/10.1371/journal.pone.0235564

Ng'wanakilala, F. (2019). *Tanzania's Mobile Phone Subscriptions Rise to Nearly 44 Million | Reuters*. Retrieved August 2, 2021, (https://www.reuters.com/article/tanzania-telecoms/tanzanias-mobile-phone-subscriptions-rise-to-nearly-44-million-idUSL3N26G26H).

NHTSA. (1990). *Motor Vehicle Fires in Traffic Crashes and the Effects of the Fuel System Integrity Standard.*

R Development Core Team. (2021). *The R Project for Statistical Computing*. Retrieved July 12, 2018, from https://www.r-project.org/

RAIS. (2021). *Road Accident Information System.*

Rezapour, M., & Ksaibati, K. (2018). Application of Multinomial and Ordinal Logistic Regression to Model Injury Severity of Truck Crashes , Using Violation and Crash Data. *Journal of Modern Transportation, 26*(4), 268-277. https://doi.org/10.1007/s40534-018-0166-x

Saleem, N., Al-bdairi, S., & Hernandez, S. (2020). International Journal of Transportation Comparison of Contributing Factors for Injury Severity of Large Truck Drivers in Run-off-Road Crashes on Rural and Urban Roadways : Accounting for Unobserved Heterogeneity. *International Journal of Transportation Science and Technology, 9*(2), 116-27. https://doi.org/10.1016/j.ijtst.2020.01.004

Salum, J. H., Kitali, A. E., Bwire, H., Sando, T., & Alluri, P. (2019). Severity of Motorcycle Crashes in Dar Es Salaam. *Traffic Injury Prevention, 20*(2), 189-195. https://doi.org/10.1080/15389588.2018.1544706

Shrestha, P. P., & Shrestha, K. J. (2017). *Factors Associated with Crash Severities in Built-up Areas along Rural Highways of Nevada: A Case Study of 11 Towns*. Elsevier Ltd. https://doi.org/10.1016/j.jtte.2016.08.003

Tu, Y. K., Kellett, M., Clerehugh, V., & Gilthorpe, M. S. (2005). Problems of Correlations between Explanatory Variables in Multiple Regression Analyses in the Dental Literature. *British Dental Journal, 199*(7), 457-461. https://doi.org/10.1038/sj.bdj.4812743

Vachal, K., Institute Upper Great Plains Transportation, Consortium Mountain-Plains, Transportation Department of, Research, &Administration Innovative Technology. (2016). Analysis of Risk Factors in Severity of Rural Truck Crashes (p. 38).

WHO. (2021). *Road Safety | WHO | Regional Office for Africa*. Retrieved July 30, 2021, from https://www.afro.who.int/health-topics/road-safety

Zhang, Z. H., He, Q., Gao, J., & Ni, M. (2018). A Deep Learning Approach for Detecting Traffic Accidents from Social Media Data. *Transportation Research Part C: Emerging*

Technologies, *86*, 580-596. https://doi.org/10.1016/j.trc.2017.11.027

Zimmerman, K., Jinadasa, D., Maegga, B., & Guerrero, A. (2015). Road Traffic Injury on Rural Roads in Tanzania: Measuring the Effectiveness of a Road Safety Program. *Traffic Injury Prevention*, *16*(5), 456-460. https://doi.org/10.1080/15389588.2014.973491

Zwerling, C., Peek-Asa, C., Whitten, P. S., Choi, S. W., Sprince, N. L., & Jones, M. P. (2005). Fatal Motor Vehicle Crashes in Rural and Urban Areas: Decomposing Rates into Contributing Factors. *Injury Prevention*, *11*(1), 24-28. https://doi.org/10.1136/ip.2004.005959

Copyright Disclaimer