Interpretable Crash Severity Prediction Models to Improve Cyclist Safety

Giuseppe Cappelli[1,2(✉)], Sofia Nardoianni[1], Mauro D'Apuzzo[1], and Vittorio Nicolosi[2]

[1] University of Cassino and Southern Lazio, Via G. Di Biasio 43, 03043 Cassino, Italy
{giuseppe.cappelli1,sofia.nardoianni,dapuzzo}@unicas.it,
cppgpp01@uniroma2.it

[2] University of Rome "Tor Vergata", Via del Politecnico, 1, 00133 Rome, Italy
nicolosi@uniroma2.it

Abstract. This study focuses on 3 years (2016–2018) of cyclist crashes in the City of Rome, Italy. As the first step, a statistical analysis was carried out. Several Cycling Crash Models were developed by using Logistic Regression Models, with a deep dive into the most influencing variables. The two proposed models at intersections and single-lane carriageways have a *McFadden* score or *pseudo-R^2* of *0.3976809* and *0.4495008*, respectively. The findings show that visibility does not play a key role in leading to a crash with a cyclist; sunny weather is positively correlated to crashes in intersections, while dry surfaces increase the chances of having crashes on single-lane carriageways, such as also the location of these roads in extra-urban environments and autumn and winter seasons. Weekdays are also related to an increase in the probability of having a crash at intersections and on single-lane carriageways. Cyclist crashes are more likely to happen in the evening and nighttime hours. Vertical and horizontal signposting decreases the probability of crashes in intersections and single-lane carriageways. High values of average daily traffic (>2000 vehicles/day) are strongly related to crashes on single-lane carriageways, and high speeds (>50 km/h) increase the probability of fatal crashes in intersections and on single-lane carriageways.

Keywords: Sustainable Mobility · Road Safety · Cyclists Safety · Cyclist Crash Severity · Cyclist Crash Models

1 Introduction

Back in 2004, on World Health Day, the *World Report on Road Traffic Injury Prevention* was launched by the World Health Organization (WHO) and the World Bank, as one of the first documents to highlight the unsustainable, unacceptable, and unsafe road traffic system [1]. As a first result of this joint action, in 2009 there was the *First Global Ministerial Conference on Road Safety* in the Russian Federation; this led to the adoption of the *Moscow Declaration*, with the declaration of *Decade of Action for Road Safety 2011–2020* [2]. The first *Global status report on road safety: Time for action* document

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
O. Gervasi et al. (Eds.): ICCSA 2025 Workshops, LNCS 15896, pp. 319–334, 2026.
https://doi.org/10.1007/978-3-031-97654-4_20

dates back to the same year [3]. In 2010, the *Decade of Action for Road Safety 2011–2020* was launched: the aim was to reverse the trend of deaths on the road [4]. The Global Plan was launched in May of the following year [5].

Ten years later, in September 2020, the UN General Assembly declared the *Decade of Action for Road Safety 2021–2030*, to reduce the number of deaths on roads by at least half and promote sustainable urban mobility, in line with the Stockholm Declaration [6, 7].

More ambitious targets are set by the European Commission, with *Vision Zero*, a strategy to reduce to zero the number of road fatalities by 2050 [8], by improving all aspects related to safe vehicles, safe infrastructure, safe road use, improved human body models, enhance automation, behavioral aspects for safer transport.

Road traffic injury is the leading cause of death among children and young people (5–29 years) and globally the number of deaths is estimated to be 1.19 million in 2021: this means that the death frequency is about 15 deaths per 100,000 population. Pedestrians represent 23% and two-wheeler users the 21% of total deaths: among the last ones, the 6% is for cyclists and the 3% for micro-mobility users [9].

Many cities in the Global North are experiencing an increase in bicycle use [10], and the problem of the increased cyclist flows may negatively influence cyclists' safety [11–14], with high exposure that is related to risk [15]. This could also not lead to a Vision Zero target [16], as prospected by the European Commission. According to the literature, the relationship between the so-called "Safety in Numbers" (e.g., the theory for that with high cyclist exposures, the cyclist safety is higher) and cyclist crashes is not well defined: a branch of literature shows a positive relationship between the number of cyclist crashes and flows [17–19], but on the other side, several studies state the opposite [20–22]. Due to this scenario, the relationships among cyclist crashes, traffic flows, and speeds are key in understanding ways to reduce the number of overall crashes by implementing interpretable and simple models to succeed in this task.

Analysing crash severity is a very studied topic both in individual and collective transport, and the issues of crash risk modelling also include the frequency of crashes and exposure issues [23, 24].

In the recent years, Machine Learning Models are widely implemented [25–28], but the main limitations and gaps related to these studies are due to the fact that such type of Models are not highly interpretable and that could be an issue when these models are related to safety aspect. Among the most interpretable models, the Logistic Regression is widely used, but, in comparisons to the previous ones, are usually less efficient.

In a paper on UK cyclist crashes [29], the authors used a machine learning algorithm called Random Forest and an econometric model called Random Parameters Logit Model to identify the characteristics of the road, environment, vehicle, driver, and cyclist that affect crash severity. Both methods identified various roadway, environmental, vehicle, driver, and cyclist-related factors associated with higher crash severity.

In the study of Engbers et al. [30], the researchers conducted a logistic regression analysis to study the relationship between personal factors and self-reported bicycle falls. The univariate models showed that age, physical and mental impairments, bicycle model, living environment, feelings of uncertainty of the cyclist, and changed cycling behavior (such as more patience and lower speed) were all related to falling off a bicycle.

While there is existing research on the causes of single bicycle crashes (SBCs), the study of Eriksonn et al. [31] seeks to examine differences in injury severity. Specifically, the study aims to investigate how injury severity is related to the characteristics of the crash and the cyclist, with a focus on SBCs. The study used binary logistic regression to analyze the odds of severe injury for different types of cyclists and situations. The authors also found that there was a higher risk of severe injury during leisure trips, on weekdays, and at intersections or road stretches compared to pedestrian and cycle paths. Furthermore, the risk of severe injury per kilometer travelled was higher for cyclists aged 45 and older during leisure trips.

Another example came from Spain [32]. A database of 2269 geo-located crashes in Madrid (involving at least one bike) was used, and this allowed the researchers to compare bike road safety between different districts in Madrid and model the severity of bike crashes using binary logit regression. The study also found that most serious crashes occurred at intersections between avenues and roads. It is important to note that, according to the authors, at intersections, cycling infrastructure is no longer protected, and the risk of a serious crash is higher.

In this paper, after a brief descriptive statistical analysis of 885 cyclist crashes in the city of Rome [33], a Logistic Regression Model is proposed with the objectives:

1. Give a general overview of the main features that influence cyclist crashes;
2. Discover the role of the vehicular speeds and the average daily traffic (ADT) in predicting the possible outcome of a crash involving cyclists and the relative probability;
3. Focus on particular crash scenarios to contextualize cyclist crashes to common patterns and improve the model prediction accuracy.
4. An interpretable and simple model that could highlight particular features to improve cyclist safety.

2 Method

2.1 Logistic Regression

Logistic Regression is a statistical model that is usually used to study the influences of independent variables on the dependent variable, which is a categorical one [34], and very often this type of model is used in classification problems [35, 36]. Linear regression, in a classification problem, may not be the best solution, so it is better to use Logistic Regression, as expressed in Eq. (1):

$$p(X) = \frac{e^{\beta_0 + \sum \beta_i X_i}}{1 + e^{\beta_0 + \sum \beta_i X_i}} \tag{1}$$

Where:

- p(X) is the predicted probability;
- xi are the independent variable or features;
- β0…βi are the regression coefficients of the model.

It is worth mentioning that this model only allows for k = 2 classes of response variables. Despite the more complex Machine Learning algorithms, such as Random Forest, Logistic Regression is characterized by a higher clarity on the interpretability side. Another advantage is their simple structure and simplicity for end users. Econometric models, on the other side, may provide unstable results due to the high amount of data that characterizes road crashes [37], so it is essential to reduce the number of variables by using feature selection algorithm.

Regarding the several features collected, to try to reduce the number of variables, they are transformed into dichotomous variables [38]. In Table 1, such type of "organization" of variables is applied also for this work. This way of pre-processing data reduces

Table 1. Description of the independent variables used for the Logistic Regression Model.

Variable	Value	Description
Location	0	In the extra-urban area
	1	Within the built-up area
Road	0	One-way road
	1	Two-way road
Stretch	0	No intersection
	1	Intersection
Surface	0	Other conditions
	1	Dry condition
Signposting	0	Other cases
	1	Vertical and horizontal
Day of the week	0	Festive
	1	Weekday
Weather conditions	0	Other cases
	1	Sunny
Hour	0	Evening or night
	1	Morning or afternoon
Season	0	Autumn or winter
	1	Spring or summer
Pavement	0	Unpaved
	1	Paved
Lighting	0	Inadequate
	1	Adequate
Visibility	0	Inadequate
	1	Adequate
Flow	0	<2000 vehicle/h
	1	>2000 vehicle/h
Speed	0	<50 km/h
	1	>50 km/h
Crash type	0	Other type of collisions
	1	Front/side collision

drastically the number of variables, reducing also the issues related to econometric models to handle high dimensional data, as highlighted by some authors [37, 39]. This selection follows the variable selection that could be noted in literature [23, 24].

3 Results

3.1 Descriptive Statistical Analysis

Road Safety, as suggested by the literature and the European and global Plans, is crucial in defining design strategies for the future and trying to reach environmental objectives. The major issues relate to urban environments, especially Metropolitan Cities worldwide. For this reason, this study focuses on 3 years (2016–2018) of cyclist crashes in the City of Rome (Italy). The Municipality of Rome provides the dataset used, in particular the *Rome Mobility Society* [33]. The dataset collects crash events for each type of road user, but the purpose here explained, the focus will be on cyclist crashes.

In the years 2016–2018, 885 crashes involving cyclists happened: 193 (21.81%) of them were unharmed, 668 (75.48%) injured, and 24 (2.71%) dead (Fig. 1).

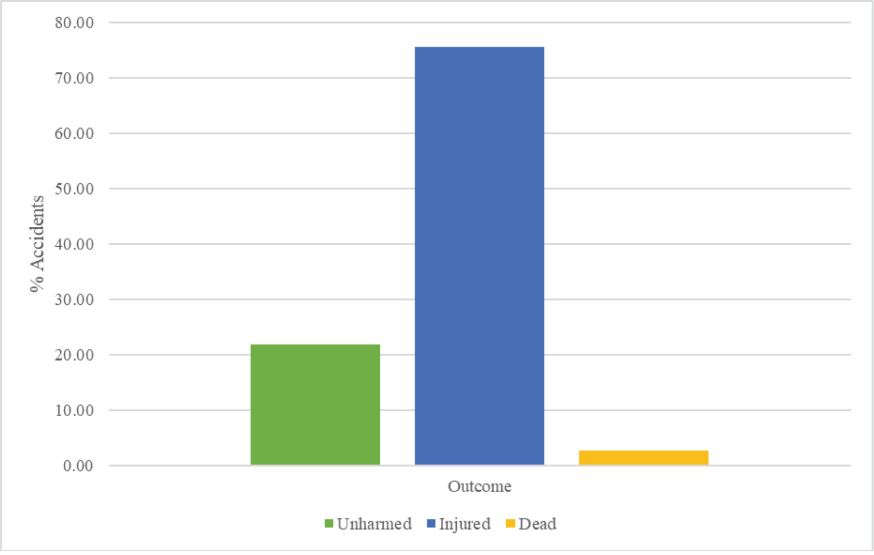

Fig. 1. Frequency of cyclist crashes according to their outcome.

The majority of crashes that lead to an injury or a fatal outcome, in frequency terms, happen in extra-urban environments; unharmed cyclists seem to be more likely in urban than extra-urban environments, as it is possible to see in Fig. 2.

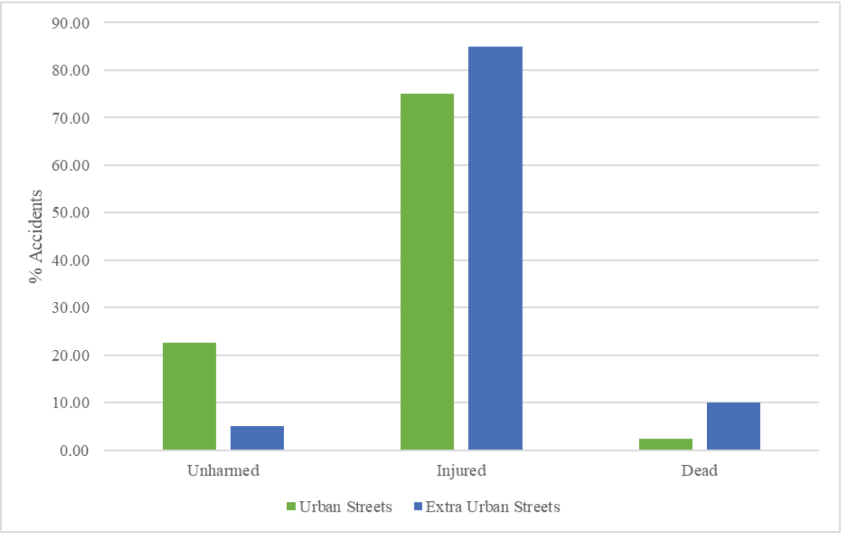

Fig. 2. Frequency of cyclist crashes according to their outcome in the urban and extra-urban environments.

In Fig. 3, the results of statistical analysis show that crashes are more likely to happen on weekdays than on weekends, and this is true also for unharmed, injured, and dead cyclist users involved in road crashes. This could be explained because during weekdays the exposure to vehicle traffic is higher than on weekends and also because more cyclists use bikes during the week.

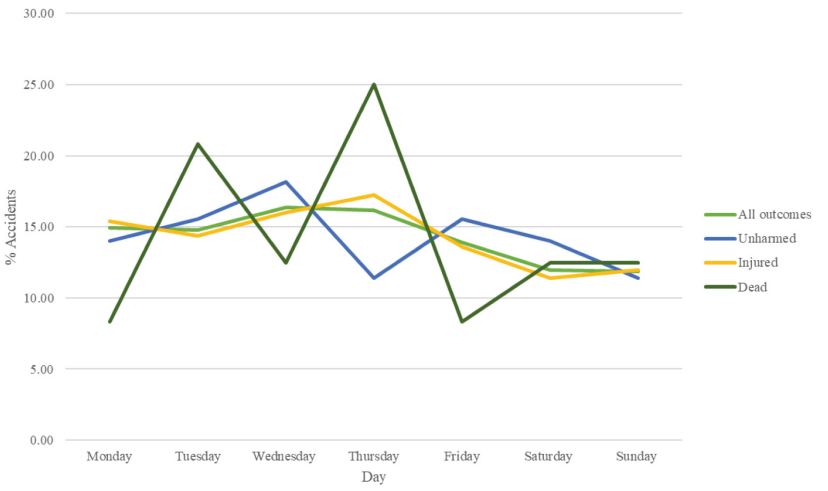

Fig. 3. Frequency of cyclist crashes according to their outcome during the week. The trend defined as “all outcome” is the sum of unharmed, injured and dead cyclists.

From a seasonal perspective, a larger number of cyclist crashes are registered in the spring and summer months than winter and autumn months (Fig. 4). The reason why this happens may be explained by the propensity of cyclists to use bikes during months when the weather is typically good. But, as it is shown from the same figure, there are more dead cyclists involved in crashes during winter months.

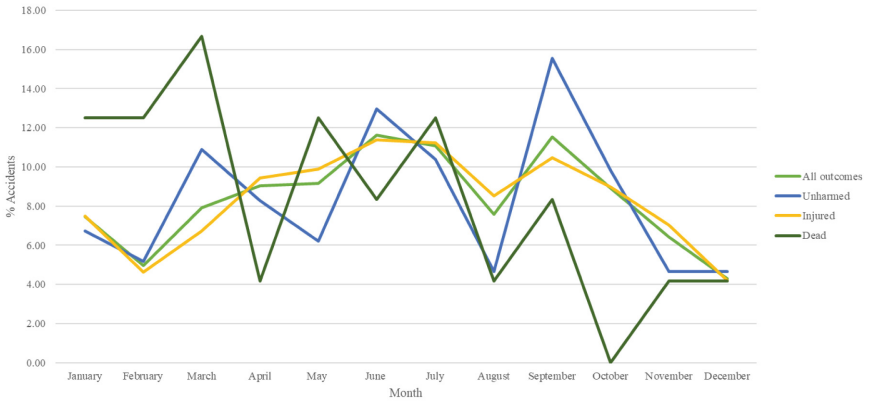

Fig. 4. Frequency of cyclist crashes according to their outcome during the years analyzed. The trend defined as "all outcome" is the sum of unharmed, injured and dead cyclists.

As shown in Fig. 5, crashes that lead to slight or non-serious outcomes follow somehow the same pattern during the day: two peaks from 10 a.m. to 12 a.m. and from 5 p.m. to 7 p.m. are related to vehicle off-peak hours; regarding the dead cyclist pattern, the peaks are reached between 8 a.m. and 10 a.m., and 7 p.m. and 9 p.m.

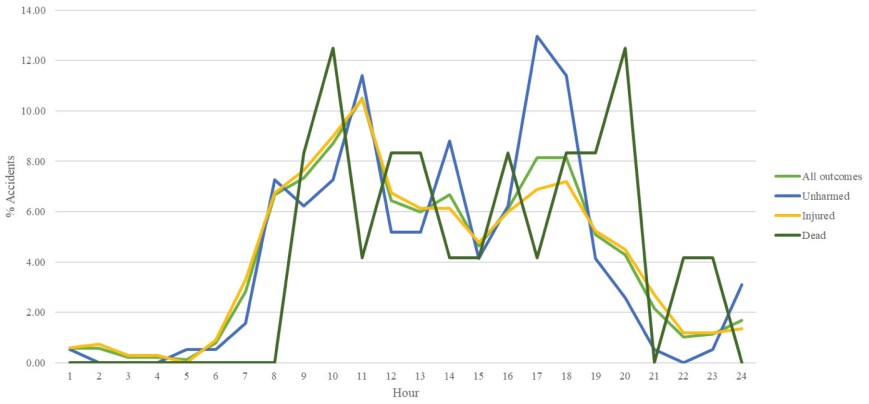

Fig. 5. Frequency of cyclist crashes according to their outcome during a day.

In Fig. 6, the relative frequency of crashes is shown: when the flow is low, there is a high likelihood that a crash leads to a fatal outcome, compared to a road with a higher traffic flow. And low flow rate increases the probability that vehicle speeds are higher: a speed higher than 30 km/h increases the probability of a fatal crash involving cyclists (Fig. 7).

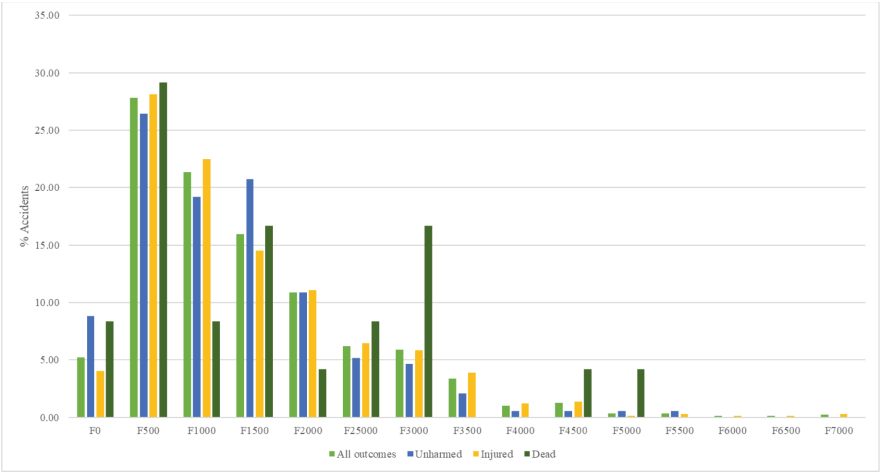

Fig. 6. Traffic flows and the likelihood of cyclist crashes.

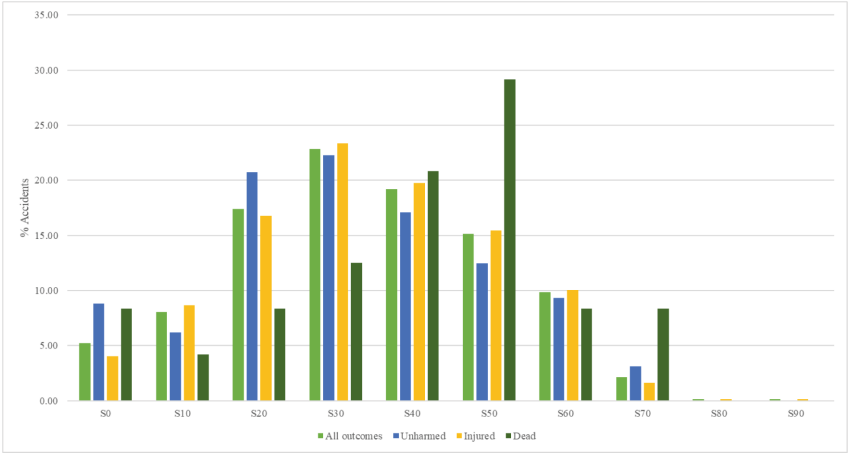

Fig. 7. The role of speed in cyclist crash.

3.2 Crash Severity Models

Cyclist crashes, but in general road crashes, are characterized by a great number of variables. Since the purpose of this work is to give shape to a simplified crash prediction

model (to be used in the real world with not-too-sophisticated tools), not-too-elaborate models such as the Logit model are used in this regard. In Table 2, a General Cyclist Crash Model and the main statistics are proposed. The model can be used to predict the probability that a cyclist crash leads to a fatal outcome, according to other several independent variables. Logistic Regression may also be useful to understand the single independent variable's role in the crash. In Table 3 the main performance metric of the proposed model is reported.

Table 2. Cyclist Crash Model for predicting the fatal outcome after a crash.

Variable	Estimate	Std. Error	z-value	Pr(>z)
Intercept	−2.29973	1.33204	−1.726	0.084263.
Location	−0.85683	0.22770	−3.763	0.000168^{***}
Day	0.24764	0.15985	1.549	0.121341
Season	−0.06881	0.15022	−0.458	0.646888
Hour	−0.43136	0.17468	−2.469	0.013534^{*}
Stretch	−0.74541	0.16530	−4.509	$6.50 \times 10^{-6***}$
Road	0.31538	0.18544	1.701	0.088995
Pavement	2.23632	0.50815	4.401	$1.08 \times 10^{-5***}$
Signposting	−0.44163	0.14739	−2.996	0.002733^{**}
Surface	0.34189	0.29826	1.146	0.251681
Weather	0.23180	0.19261	1.203	0.228810
Visibility	1.68029	1.21070	1.388	0.165180
Lighting	−1.28513	0.47312	−2.716	0.006602^{**}
Crash type	0.09009	0.16609	0.542	0.587539
ADT	−0.23563	0.15217	−1.548	0.121513
Speed	0.77232	0.14980	5.156	$2.53 \times 10^{-7***}$

Table 3. Performance metrics and main statistics for the three Cyclist Crash Models analyzed.

Estimator	Fatal
McFadden (pseudo-R^2)	0.1323
Akaike information criterion (AIC)	882.22
X^2	129.64
Training accuracy	0.6539
Test accuracy	0.6667
Training F1 score	0.6772
Test F1 score	0.6667
p-value	0.000001

To improve the prediction accuracy of the proposed model, two particular scenarios are selected, and in particular, crashes at intersections and on single-lane carriageways. As shown in Table 4, the prediction performances of these two scenario-based models,

according to the McFadden metric (but also accuracy and F1 score), have improved by drawing these sub-sets to study some crash scenarios: it is worth mentioning these models that have a McFadden higher than 0.2 (which is considered a lower boundary for good performances [40]). Tables 5 and 6 summarize the proposed models and other important metrics, such as the p-value, at intersections and single-lane carriageways.

Table 4. Performance metrics and main statistics for the two Cyclist Crash Models analyzed at the intersection and Single-Lane Carriageway.

Estimator	Intersection	Single-lane Carriageway
McFadden (pseudo-R^2)	0.4034	0.4450
Akaike information criterion (AIC)	244.81	156.16
X^2	145.24	101.17
Training accuracy	0.8192	0.8293
Test accuracy	0.7969	0.7317
Training F1 score	0.8239	0.8292
Test F1 score	0.8000	0.7556
p-value	0.000001	0.000001

Table 5. Cyclist Crash Model for predicting the fatal outcome after an crash, when a collision happens at intersections. Significance codes: 0 '***'; 0.001 '**'; 0.01 '*'; 0.05 '.'; 0.1 ' '.

Variable	Estimate	Std. Error	z-value	Pr(>z)
Intercept	−16.5769	5.2803	−3.139	0.001693^{**}
Location	2.8536	3.1374	0.910	0.363066
Day	2.6638	0.6406	4.158	$3.21 \times 10^{-5***}$
Season	0.2430	0.3163	0.768	0.442330
Hour	−0.7368	0.3089	−2.385	0.017064^{*}
Road	0.3320	0.3186	1.042	0.297251
Pavement	3.6950	1.1724	3.152	0.001624^{**}
Signposting	−1.5566	0.3582	−4.345	$1.39 \times 10^{-5***}$
Surface	0.3554	1.0446	0.340	0.733655
Weather	2.3712	0.6632	3.575	0.000350^{***}
Visibility	4.6682	1.7771	2.627	0.008616^{**}
Lighting	1.4261	3.1553	0.452	0.651296
Crash type	−1.1237	0.3337	−3.367	0.000759^{***}
ADT	−0.6092	0.3706	−1.644	0.100221
Speed	2.5326	0.4286	5.909	$3.44 \times 10^{-9***}$

Table 6. Cyclist Crash Model for predicting the fatal outcome after an crash, when a collision happens on a single-lane carriageway.

Variable	Estimate	Std. Error	z-value	Pr(>z)
Intercept	−15.2634	5.0675	−3.012	0.002595**
Location	−1.9088	0.5530	−3.452	0.000557***
Day	4.3735	1.0547	4.147	3.37×10^{-5}***
Season	−0.7358	0.3742	−1.967	0.049235*
Hour	−0.7170	0.4193	−1.710	0.087254.
Stretch	0.1035	0.4088	0.253	0.800053
Pavement	2.0296	1.2035	1.686	0.091711.
Signposting	−1.6981	0.5165	−3.287	0.001011**
Surface	5.7478	2.1237	2.707	0.006798**
Weather	0.1640	0.6900	0.238	0.812155
Visibility	3.4075	3.5376	0.963	0.335425
Lighting	2.5496	3.3307	0.765	0.443981
Crash type	1.0891	0.4598	2.369	0.017853*
ADT	1.2271	0.5207	2.356	0.018449*
Speed	1.6448	0.4622	3.559	0.000373***

In addition, a sensitivity analysis is also carried out to quantify the weight of vehicular speed and ADT in predicting the severity of a cyclist crash. In Table 7 a summary is proposed. As it possible to observe, when the variable *Speed* is omitted, the performances of the models in terms of McFadden decrease in a higher way than when the variable *ADT* is removed from the model. This also highlights the great role played by vehicular speeds in crash severity outcomes.

Table 7. Summary of the sensitivity analysis performed: when the variable Speed is removed from the models (M_1), the McFadden decreases at a high rate. When *ADT* is removed (M_2), there is also a decrease in term of McFadden, but lower than what was seen in the previous model (M_1).

Estimator	Intersection	Single-lane Carriageway
M_0 (with all variables)	0.4034	0.4495
M_1 (without *speed*)	0.2571	0.3784
M_2 (without *ADT*)	0.4024	0.4253
M_3 (without *speed* and *ADT*)	0.2552	0.3588

4 Discussion and Conclusions

As shown earlier in Table 3, General Crash Models seem to have not good prediction performances (as shown in Table 4). Despite this issue, with these types of models, it is possible to have an insight into which variables could be not negligible in cyclist crashes and their respective outcomes.

When cyclists die after a crash, more variables influence this type of outcome. According to the model in Table 3, extra-urban areas are crucial to define this outcome, usually on a weekday, during evening and night hours, where there are no intersections, on paved two-way roads, without vertical and horizontal signposting, under poor lighting conditions and with elevated values of vehicle speeds.

As shown in Table 4, the performance metrics of these types of models are not so valuable. For this reason, this study makes efforts to find some commonalities among road crashes to better find a valuable relationship among variables and improve prediction performances. Crash nature, road characteristics, and road types seem to be common factors characterizing cyclist crashes, but also "environmental" aspects that are common features in small and large cities, in Europe and worldwide.

In the proximity of intersections, when neither vertical nor horizontal signposting is available, during night hours on weekdays, with speeds above 50 km/h on paved roads, and with sunny weather, fatal crashes are more likely to happen (Table 5).

Regarding the last model analyzed, when a collision happens on a single-lane carriageway, extra-urban location, weekdays, autumn and winter seasons, night hours, paved and dried road without vertical and horizontal signposting, average daily traffic above 2000 vehicles/hour and speed above 50 km/h, increase the chances of fatal crashes (as it is possible to see in Table 6).

Interesting findings come out when comparing the different variables, and their role within the models, of the two scenario-based crash models to predict the outcome of the crash. Visibility does not play a key role in leading to a crash with a cyclist; sunny weather is positively correlated to crashes in intersections, while dry surfaces increase the chances of having crashes on single-lane carriageways, such as also the location of these roads in extra-urban environments and autumn and winter seasons. Weekdays are also related to an increase in the probability of having a crash at intersections, and single-lane carriageways. Cyclist crashes are more likely to happen in the evening and nighttime hours. Vertical and horizontal signposting decreases the probability of crashes in intersections and single-lane carriageways. High values of average daily traffic (>2000 vehicles/day) are strongly related to crashes on single-lane carriageways, and high speeds (> 50 km/h) increase the probability of fatal crashes in intersections and on single-lane carriageways.

The performances obtained with the simple Logistic Crash Models for Intersections and Single Lane Carriageway (see Table 4) are comparable also with more advantage Logistic Models. In the study of Das et al. [41], the McFadden associated with the Logit model proposed in that paper is 0.084 (lower than the pseudo-R^2 in this work), while the Mixed Logit Model reaches a McFadden of 0.4161 (comparable with the simple Logit Models here proposed). It is worth mentioning also other studies that implement Mixed Logit Models. For instance, in the work of Ye et al. [42], the authors study vehicle-electric bicycle crashes and the model proposed reached a McFadden of 0.5090, while in the

studies of Chen et al. [43] and Wang et al. [44], the models present a McFadden of 0.06 and 0.315, respectively. In the recent work of Scarano et al. [29], the authors proposed a Mixed Logit Model that present a McFadden of 0.21 and a F1score of 0.72, but also a Random Forest with a F1 score of 0.88 (comparable with the F1 score obtained with the simple Logit Model, and in particular 0.80 and 0.75 for Intersection and Single Lane Carriageway). In summary, the performances of the models here proposed to predict cyclist crashes are comparable with high performances models, such as Mixed Logit and Random Forest Models.

As a positive aspect of Logistic Regression, their interpretability could give information on how the single variables could increase or decrease the probability of having non-serious, serious, and fatal crashes and so understand how to address some issues.

The interpretability side opens the doors also to relevant findings and practical implications, especially for the practice of urban planning.

It is worth mentioning also the limitations of this study. The first one is related to the few data available (only for 3 years) and the dataset does not contain information on gender and age of road users involved in the crash. For these reasons, not all the context categories of type of crashes, type, and road characteristics are analyzed, and some of them lead to poor results. It is also advisable to study other cities to highlight the same or different characteristics.

What are the main implications of this study and new avenues for future research? First the most important implication of this study is related to the interpretability of the model combined with good prediction performances that could be compared with performances of more sophisticated models (and this aspect wants also to close the research gap that is an emerging trend, i.e., the huge implementation of Machine Learning Algorithms that are not so interpretable as Econometric Models). Secondly, simple models could be implemented easily to predict the severity of the crashes and lastly, the scenario-based approach is useful not only to improve model performances, but also to highlight influencing feature related to that particular scenario.

As future research developments, could be interesting to collect more recent data on the city of Rome, but also on other cities; enlarge the methodology to other road users, and use also different type of dataset: in general, this research, is a preliminary step in defining Cyclist Crash Models for predicting the probability that a crash could happen, based on road and users' characteristics.

Acknowledgments. The research reported in this paper was developed in the EMOTIVEs Research Project, supported by the "FESR Lazio 2021-2027 Program - RSI Competitive Repositioning Notice".

Author Contributions. **Conceptualization**, C.G. and N.S.; **methodology**, C.G. and N.S.; **software**, C.G. and N.S.; **validation**, C.G. and N.S.; **formal analysis**, C.G. and N.S.; **investigation**, C.G. and N.S.; **resources**, C.G. and N.S.; **data curation**, C.G. and N.S.; **writing—original draft preparation**, C.G. and N.S.; **writing—review and editing**, C.G. and N.S.; **visualization**, C.G. and N.S..; **supervision**, D.A.M. and N.V.; **project administration**, D.A.M. and N.V.; **funding acquisition**, D.A.M. and N.V. All authors have read and agreed to the published version of the manuscript.

References

1. WHO, World Health Organization: World Report on Road Traffic Injury Prevention. Available online: https://www.who.int/southeastasia/publications/i/item/world-report-on-road-traffic-injury-prevention, last accessed 16 Apr 2024
2. WHO, World Health Organization: First Global Ministerial Conference on Road Safety. Available online: https://www.who.int/publications/m/item/first-global-ministerial-conference-on-road-safety, last accessed 16 Apr 2024
3. WHO, World Health Organization: Global status report on road safety: Time for action. Available online: https://www.who.int/southeastasia/publications/i/item/9789241563840, last accessed 16 Apr 2024
4. WHO, World Health Organization: Decade of Action for Road Safety 2011–2020. Available online: https://www.who.int/groups/united-nations-road-safety-collaboration/decade-of-action-for-road-safety-2011-2020, last accessed 16 Apr 2024
5. WHO, World Health Organization: Global Plan for the Decade of Action for Road Safety 2011–2020. Available online: https://www.who.int/publications/m/item/global-plan-for-the-decade-of-action-for-road-safety-2011-2020, last accessed 16 Apr 2024
6. WHO, World Health Organization: Decade of Action for Road Safety 2021–2030. Available online: https://www.who.int/teams/social-determinants-of-health/safety-and-mobility/decade-of-action-for-road-safety-2021-2030, last accessed 16 Apr 2024
7. UN, United Nations: Conference on the Human Environment in Stockholm. Available online: https://www.un.org/en/conferences/environment/stockholm1972, last accessed 16 Apr 2024
8. European Commission, Road Safety Plan Vision Zero. Available online: https://road-safety.transport.ec.europa.eu/index_en, last accessed 16 Apr 2024
9. WHO, World Health Organization: Global status report on road safety 2023. Available online: https://www.who.int/teams/social-determinants-of-health/safety-and-mobility/global-status-report-on-road-safety-2023, last accessed 16 Apr 2024
10. Schepers, P., Helbich, M., Hagenzieker, M., de Geus, B., Dozza, M., Agerholm, N., et al.: The development of cycling in European countries since 1990. Eur. J. Transp. Infrastruct. Res. **21**(2), 41–70 (2021)
11. Schepers, P., Stipdonk, H., Methorst, R., Olivier, J.: Bicycle fatalities: trends in crashes with and without motor vehicles in The Netherlands. Transport. Res. F: Traffic Psychol. Behav. **46**, 491–499 (2017)
12. D'Apuzzo, M., Cappelli, G., Nardoianni, S., Nicolosi, V., Evangelisti, A.: A preliminary effort to develop a framework of distance decay functions for new urban active mobility. In: International Conference on Computational Science and Its Applications, pp. 194–208. Springer Nature Switzerland, Cham (2023, June)
13. Cappelli, G., D'Apuzzo, M., Nardoianni, S., Nicolosi, V.: Exploring the influences of safety and energy expenditure parameters on cycling. Sustain. For. **16**(7), 2739 (2024)
14. D'Apuzzo, M., Evangelisti, A., Cappelli, G., Nicolosi, V.: An introductory step to develop distance decay functions in the Italian context to assess the modal split to e-bike and e-scooter. In: In 2022 Second International Conference on Sustainable Mobility Applications, Renewables and Technology (SMART), pp. 1–8. IEEE (2022, November)
15. Uijtdewilligen, T., Ulak, M.B., Wijlhuizen, G.J., Bijleveld, F., Dijkstra, A., Geurs, K.T.: How does hourly variation in exposure to cyclists and motorised vehicles affect cyclist safety? A case study from a Dutch cycling capital. Saf. Sci. **152**, 105740 (2022)
16. Wegman, F., Schepers, P.: Safe system approach for cyclists in The Netherlands: towards zero fatalities and serious injuries? Accid. Anal. Prev. **195**, 107396 (2024)
17. Tasic, I., Elvik, R., Brewer, S.: Exploring the safety in numbers effect for vulnerable road users on a macroscopic scale. Accid. Anal. Prev. **109**, 36–46 (2017). https://doi.org/10.1016/j.aap.2017.07.029. Epub 2017 Oct 10

18. Fyhri, A., Sundfør, H.B., Bjørnskau, T., Laureshyn, A.: Safety in numbers for cyclists-conclusions from a multidisciplinary study of seasonal change in interplay and conflicts. Accid. Anal. Prev. **105**, 124–133 (2017). https://doi.org/10.1016/j.aap.2016.04.039. Epub 2016 May 28
19. Jacobsen, P.L.: Safety in numbers: more walkers and bicyclists, safer walking and bicycling. Inj. Prev. **9**(3), 205–209 (2003). https://doi.org/10.1136/ip.9.3.205. Erratum in: Inj. Prev. 2004 Apr;10(2):127. PMID: 12966006; PMCID: PMC1731007
20. Jacobsen, P.L., Ragland, D.R., Komanoff, C.: Safety in numbers for walkers and bicyclists: exploring the mechanisms. Inj. Prev. **21**(4), 217–220 (2015)
21. Lee, J., Abdel-Aty, M., Xu, P., Gong, Y.: Is the safety-in-numbers effect still observed in areas with low pedestrian activities? A case study of a suburban area in the United States. Accid. Anal. Prev. **125**, 116–123 (2019). https://doi.org/10.1016/j.aap.2019.01.037. Epub 2019 Feb 7
22. Jacobsen, P.L.: Safety in numbers: more walkers and bicyclists, safer walking and bicycling. Inj. Prev. **21**(4), 271–275 (2015)
23. Bonera, M., Barabino, B., Yannis, G., Maternini, G.: Network-wide road crash risk screening: a new framework. Accid. Anal. Prev. **199**, 107502 (2024)
24. Barabino, B., Bonera, M., Maternini, G., Porcu, F., Ventura, R.: Refining a crash risk framework for urban bus safety assessment: evidence from Sardinia (Italy). Reliab. Eng. Syst. Saf. **245**, 110003 (2024)
25. Komol, M.M.R., Hasan, M.M., Elhenawy, M., Yasmin, S., Masoud, M., Rakotonirainy, A.: Crash severity analysis of vulnerable road users using machine learning. PLoS One. **16**(8), e0255828 (2021)
26. Janstrup, K.H., Kostic, B., Møller, M., Rodrigues, F., Borysov, S., Pereira, F.C.: Predicting injury-severity for cyclist crashes using natural language processing and neural network modelling. Saf. Sci. **164**, 106153 (2023)
27. Scarano, A., Aria, M., Mauriello, F., Riccardi, M.R., Montella, A.: Systematic literature review of 10 years of cyclist safety research. Accid. Anal. Prev. **184**, 106996 (2023)
28. Birfir, S., Elalouf, A., Rosenbloom, T.: Building machine-learning models for reducing the severity of bicyclist road traffic injuries. Transp. Eng. **12**, 100179 (2023)
29. Scarano, A., Riccardi, M.R., Mauriello, F., D'Agostino, C., Pasquino, N., Montella, A.: Injury severity prediction of cyclist crashes using random forests and random parameters logit models. Accid. Anal. Prev. **192**, 107275 (2023)
30. Engbers, C., Dubbeldam, R., Brusse-Keizer, M.G.J., Buurke, J.H., De Waard, D., Rietman, J.S.: Characteristics of older cyclists (65+) and factors associated with self-reported cycling accidents in The Netherlands. Transport. Res. F: Traffic Psychol. Behav. **56**, 522–530 (2018)
31. Eriksson, J., Niska, A., Forsman, Å.: Injured cyclists with focus on single-bicycle crashes and differences in injury severity in Sweden. Accid. Anal. Prev. **165**, 106510 (2022)
32. Guirao, B., Gálvez-Pérez, D., Casado-Sanz, N.: The impact of the cyclist infrastructure type on bike accidents: the experience of Madrid. Transp. Res. Procedia. **71**, 403–410 (2023)
33. Roma Mobilità: Available online: https://romamobilita.it/it, last accessed 22 Apr 2024
34. Nick, T.G., Campbell, K. M.: Logistic regression. Topics in biostatistics, 273–301 (2007).
35. James, G., Witten, D., Hastie, T., Tibshirani, R.: An Introduction to Statistical Learning, vol. 112, p. 138. springer, New York, NY (2013)
36. Hosmer Jr., D.W., Lemeshow, S., Sturdivant, R.X.: Applied Logistic Regression. John Wiley & Sons (2013)
37. Mannering, F., Bhat, C.R., Shankar, V., Abdel-Aty, M.: Big data, traditional data and the tradeoffs between prediction and causality in highway-safety analysis. Anal. Methods Accid. Res. **25**, 100113 (2020)
38. Eboli, L., Forciniti, C., Mazzulla, G.: Factors influencing accident severity: an analysis by road accident type. Transp. Res. Procedia. **47**, 449–456 (2020)

39. Meuleners, L.B., Fraser, M., Johnson, M., Stevenson, M., Rose, G., Oxley, J.: Characteristics of the road infrastructure and injurious cyclist crashes resulting in a hospitalisation. Accid. Anal. Prev. **136**, 105407 (2020)
40. Domencich, T.A., McFadden, D.: Urban travel demand-a behavioral analysis (No. Monograph) (1975).
41. Das, S., Tamakloe, R., Zubaidi, H., Obaid, I., Rahman, M.A.: Bicyclist injury severity classification using a random parameter logit model. Int. J. Transp. Sci. Technol. **12**(4), 1093–1108 (2023)
42. Ye, F., Wang, C., Cheng, W., Liu, H.: Exploring factors associated with cyclist injury severity in vehicle-electric bicycle crashes based on a random parameter logit model. J. Adv. Transp. **2021**(1), 5563704–5563712 (2021)
43. Chen, C., Anderson, J.C., Wang, H., Wang, Y., Vogt, R., Hernandez, S.: How bicycle level of traffic stress correlate with reported cyclist accidents injury severities: a geospatial and mixed logit analysis. Accid. Anal. Prev. **108**, 234–244 (2017)
44. Wang, T., Chen, J., Wang, C., Ye, X.: Understand e-bicyclist safety in China: crash severity modeling using a generalized ordered logit model. Adv. Mech. Eng. **10**(6), 1687814018781625 (2018)