Heliyon 9 (2023) e23062

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Analysis and prediction of injury severity in single micromobility crashes with Random Forest

Almudena Sanjurjo-de-No [a], Ana María Pérez-Zuriaga [b,*], Alfredo García [b]

[a] Universidad Politécnica de Madrid (UPM), Ramiro de Maeztu, 7, 28040, Madrid, Spain
[b] Highway Engineering Research Group (HERG), Universitat Politècnica de València, Camino de Vera, s/n, 46022 Valencia, Spain

ARTICLE INFO

Keywords:
Injury severity
Micromobility
Random Forest
Road safety
Urban area

ABSTRACT

Urban micromobility represents a significant shift towards sustainable cities, underscoring the paramount importance of its safety. With the surge in micromobility adoption, collisions involving micromobility devices, such as bicycles and e-scooters, have surged in recent years. The second most common crash type involving these vehicles is one that only involves a micromobility vehicle (single micromobility crashes). This study analyzed 6030 single micromobility crashes that occurred in Spanish urban areas from 2016 to 2020. The Random Forest methodology was applied to create a classification model for the purpose of characterizing these crashes, predicting their injury severity, and identifying the primary influencing factors. To address the issue of imbalanced data, resulting from the relatively smaller dataset of fatal and seriously injured crashes compared to slightly injured ones, the Synthetic Minority Oversampling Technique (SMOTE) was applied.

The results indicate that certain behaviors, such as not wearing a helmet, riding for leisure, and instances of speeding violations, have the potential to increase injury severity. Additionally, crashes occurring at intersections or at cycle lanes with bad pavement conditions are likely to result in more severe outcomes. Furthermore, the concurrent presence of various other factors also contributes to an escalation in crash injury severity.

These findings have the potential to provide valuable insights to authorities, assisting them in the decision-making process to enhance micromobility safety and thereby promoting the creation of more equitable and sustainable urban environments.

1. Introduction

Urban mobility patterns have changed in recent years worldwide thanks to the strong rise of micromobility. Micromobility refers to the use of microvehicles, defined as vehicles with a mass not exceeding 350 kg and a design speed no greater than 45 km/h [1,2]. Micromobility vehicles (MMV) include bicycles and Personal Mobility Devices (PMDs), which encompass stand-up e-scooters (e-scooters), as well as other PMDs like Segways, electric- or e-skateboards and self-balancing motorized scooters (hoverboards). Among these micromobility vehicles, bicycles constitute the most widely utilized mode of transportation, followed by e-scooters, which have gained a lot of popularity in recent years [3,4]. These micromobility vehicles have a direct impact, particularly within urban area, where they may interact with both pedestrians and motor vehicles [5,6].

* Corresponding author.
E-mail address: anpezu@tra.upv.es (A.M. Pérez-Zuriaga).

https://doi.org/10.1016/j.heliyon.2023.e23062
Received 20 April 2023; Received in revised form 24 November 2023; Accepted 24 November 2023
Available online 30 November 2023

The significant increase in micromobility vehicles use in recent years has also led to a growing in the number of crashes involving these vehicles [6–9]. According to the European Transport Safety Council [10], 53 % of all cyclist fatalities within the European Union result from collisions with passenger cars and 16 % of cyclists fatalities occur in single-bicycle crashes (SBCs). This last percentage is higher than the proportion of cyclist fatalities resulting from collisions with heavy goods vehicles (13 %), vans (7 %), buses (2 %), and other vehicles (6 %). Although most cyclist fatalities arise from crashes involving motor vehicles and bicycles, it is important to note that a significant number of emergency department attendances are from SBCs [11], which also constitute a major contributor to severe cyclist injuries [12]. Additionally, the crash frequency is expected to be even higher for SBCs, especially when considering the substantial underreporting of data in these incidents compared to cases involving motor vehicles [13]. Therefore, understanding the causes and factors influencing the occurrence and severity of such crashes is relevant to develop effective countermeasures.

Several studies have focused on characterizing SBCs and identifying contributory factors. Two main types of SBCs have been identified: (i) falls from a bicycle and (ii) collisions between cyclists and objects [14]. The factors contributing to these incidents encompass the behavior or presence of other users, cyclist distraction and behavior, bicycle saddle height, and the width of the cycling facilities. However, the prevalence of these factors may depend on the region under study and the season. For instance, in Sweden [15,16] and Denmark [17], the most common type of crash was skidding on ice or snow, adding an environmental dimension to the contributing factors. Additionally, the slippery road conditions resulting from snow or ice emerge as an additional contributory factor in these crashes. Furthermore, engaging in cycling after the consumption of alcohol or drugs has been associated with a higher risk of SBCs [18].

Injuries stemming from crashes may vary depending on the crash type, with hip and upper leg, as well as shoulder and upper arm injuries, being the most frequently reported [15]. Moreover, several attributes may be related to injury severity [16], including the cyclist gender and age, the purpose of the trip, crash location, helmet usage, and whether the crash occurs on a weekday or weekend.

Considering the distinctions between electric bicycles and traditional bicycles, an examination of their safety profiles has been conducted [19]. After adjusting for age, gender and bicycle usage frequency, electric vehicle users are found to have a slightly higher likelihood of being involved in crashes requiring treatment. These differences are small, especially where the speeds of electric bicycles and traditional bicycles are similar. Contributory factors to the occurrence of SBC are also quite similar across both types of bicycles [20].

Regarding single e-scooter crashes, the studies are more limited, and their safety is less-understood compared to other transport means [21]. Falls are the most common type of single e-scooter crash and account for the majority of associated injuries [5,8,9,22,23]. Moreover, falls are in more than 80 % of these crashes which reports injuries, following by collision with objects (11.0 %) and being hit by moving vehicles or objects (8.8 %) [9]. Head injuries are the most frequently observed consequence of single e-scooter crashes, accompanied by injuries to the upper and lower extremities of the body [5,6,8,9,22,24,25].

Crash injury severity was analyzed in several studies using different methodologies. Statistical modeling, a conventional approach, has been applied in crash severity analysis for a considerable period due to its capability to provide reliable insights into crash likelihood with easily interpretable results [26]. However, statistical modeling requires certain assumptions about the underlying data distribution and predefined relationships between dependent and independent variables. In contrast, machine learning techniques do not rely on pre-assumed relationships between variables, contributing to their growing adoption in this research domain. Among these techniques, Random Forest, Support Vector Machine, and Decision Tree analytical methods presented have consistently demonstrated superior performance in multiple studies [26]. Moreover, in the comparative analysis of Random Forest and Classification and Regression Tree (CART) for identifying significant variables linked to injury severity categories and for classifying and predicting vehicle drivers' injury severity, Random Forest achieved better accuracy compared to CART [27,28].

Although some researchers have analyzed SBCs and single e-scooter crashes and the injuries resulting from these incidents, there are still many unexplored aspects to comprehensively understand these crashes [13], especially considering that most of the studies are based on a limited sample of crashes.

Therefore, in this research, an in-depth study of the factors that can influence the severity of injuries resulting from a single micromobility crash has been conducted. The study not only encompasses single bicycle crashes but also other single crashes involving a PMD, especially e-scooters. Factors related to rider, micromobility vehicle, crash, and infrastructure have been considered in the classification and predictive model developed using the Random Forest methodology.

The rest of this paper is structured as follows: Section 2 introduces the crash database and outlines the methodology used for this research; Section 3 presents all the results and provides a discussion of the findings; and Section 4 offers the conclusions of this research and outlines directions for future works.

2. Material and methods

2.1. Crash data

This study is based on crashes that occurred in Spanish urban areas from 2016 to 2020, involving only one micromobility user. These crashes records are included in the National Registry of Traffic Accidents victims provided by the Spanish General Directorate of Traffic (Dirección General de Tráfico, DGT). This database is derived from police records.

The initial five-year database included 305,689 crashes that took place in urban areas during the analyzed period. Among these, 29,913 incidents involve at least one micromobility user (bicycle, e-scooter, or other PMD).

During this period, bicycles involved in crashes have remained relatively stable. However, the number of PMDs (mainly e-scooters) has significantly increased in recent years, particularly since 2018. Consequently, the number of micromobility crashes has also shown a significant increase, as depicted in Fig. 1.

Considering the study objective, collisions involving motor vehicles, multiple micromobility users, or pedestrians were excluded from the database. Among all the crashes involving at least one micromobility user, 6030 are categorized as single micromobility crashes. These crashes are defined as incidents where only one micromobility user is involved, with no interaction with motor vehicles, other micromobility vehicles (neither bicycle nor PMD), or pedestrians. The proportion of crashes involving only one micromobility vehicle has consistently increased, rising from 16.91 % in 2016 to 26.51 % in 2020, as illustrated in Fig. 2.

More than half of all reported single micromobility crashes are attributed to falls (66.02 %), followed by overturning crashes (21.87 %), and a smaller proportion involving obstacle collisions (9.64 %). Other reported single micromobility crashes include collisions with animals, run-off crashes, and falling off cliffs (Fig. 3). Approximately 10 % of all these incidents result in severe injuries or fatalities.

In this research, only crashes involving a single micromobility vehicle have been analyzed. While these crashes are less frequent and typically less severe than collisions between micromobility vehicles and motor vehicles (Fig. 2), they have significant implications for micromobility users. Furthermore, their frequency has increased in recent years, and there is still a lack of knowledge in the literature regarding this type of collision.

The provided database includes information about the crashes, vehicles, and riders. Several variables were retrieved from the database and reorganized for inclusion in the developed model, as indicating in the following subsections. All variables have been categorized in a binary manner, whith 0 denoting the absence of the variable and 1 indicating its presence. The only variables where the values 0 and 1 denote something different from what was previously stated are *WEEKEND*, *GENDER*, and *BICYCLE*. Each of these variables will be described in detail in the following subsections, explaining the meaning of their values.

Multicollinearity among all factors was analyzed using the Pearson correlation coefficient. High collinearity is considered to exist when this coefficient is, in absolute value, greater than 0.7 [29]. The highest correlation among all variables is −0.67. Therefore, all identified correlations are lower, in absolute value, than 0.7, so all these variables were included in the model. Additionally, one of the advantages of the Random Forest methodology is that it is more robust to multicollinearity compared to other traditional statistical methodologies [30].

2.1.1. Crash information

Data from the DGT crash database were extracted to define seven variables for inclusion in the model (Table 1).

Crash injury severity (*SEVERITY* variable) was selected as the response variable and categorized into “slightly injured” (value 0) and “fatal and seriously injured” (value 1). The "slightly injured" category includes micromobility users who have sustained minor injuries due to collisions involving micromobility vehicles. In contrast, the "fatal and seriously injured" group refers to micromobility users who have suffered severe injuries, requiring hospitalization exceeding 24 h, or fatalities resulting from such collisions.

The *SEVERITY* variable faces an issue of imbalance due to a significantly lower number of fatalities and serious injuries (598 crashes) compared to minor injuries (5432 crashes).

Regarding the date, the variable *WEEKEND* was defined based on the variables from the original database indicating the day, month, and year of the collision. The *WEEKEND* variable takes the value 0 if the crash occurred on a weekday and the value 1 for those occurred on weekends.

Time information was used to define two variables: *MORNING*, covering crashes occurring from sunrise to 2 p.m., and *AFTERNOON,* from 3 p.m. to sunset. Both variables are assigned a value of 0 when the crash occurs at night.

Finally, three variables were established to define the type of crash: *OBSTACLE_COLLISION*, *FALL*, and *OVERTURNING*. Other types of crashes (collisions with animals, run-off crashes, and falling off cliffs) are represented in the model when all three variables are set to 0.

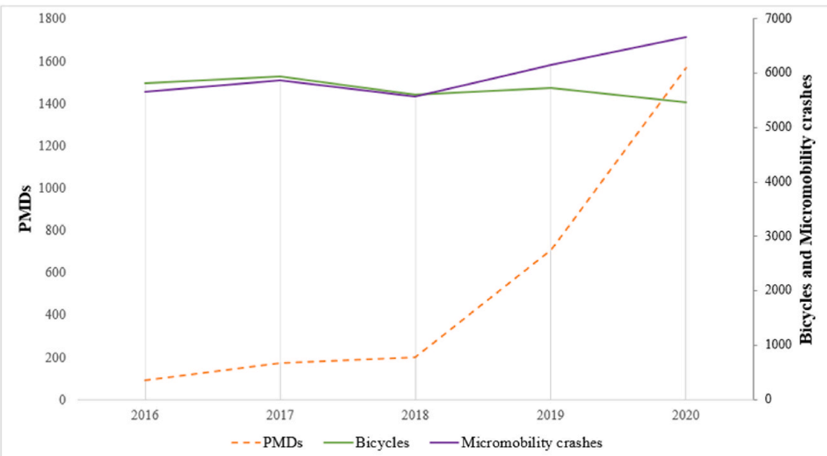

Fig. 1. Evolution of micromobility crashes and users in urban areas in Spain (2016–2020).

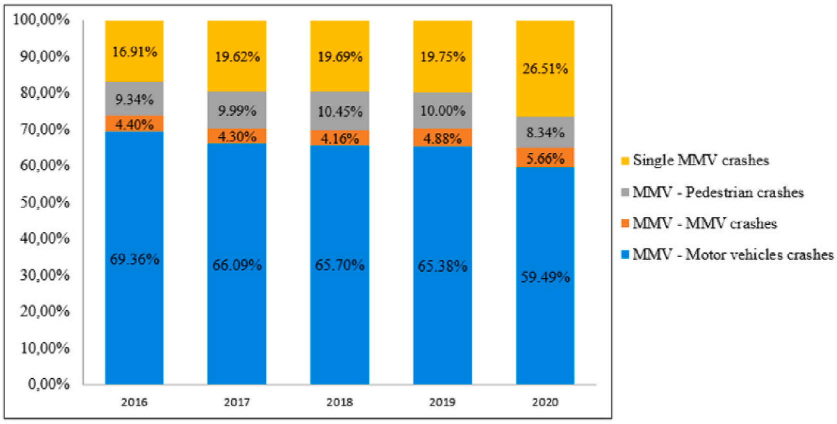

Fig. 2. Distribution of micromobility crashes in urban area in Spain according to users involved (2016–2020).

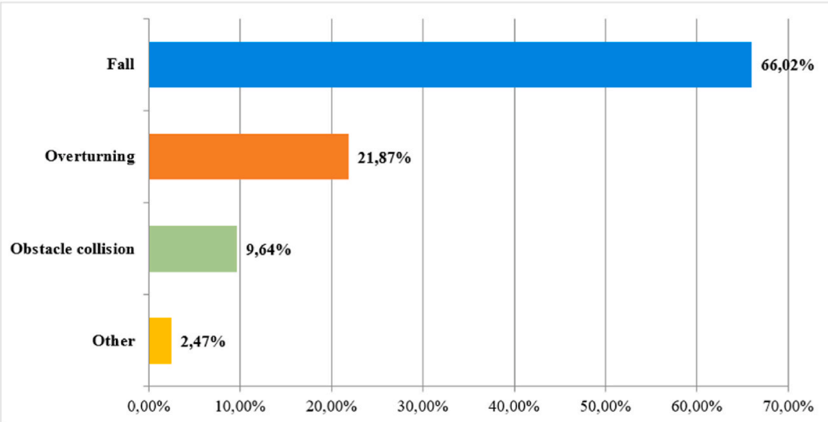

Fig. 3. Single micromobility crash types in urban areas in Spain (2016–2020).

Table 1
Crash information variables.

Variables		Total (%)	Micromobility crashes	
			Fatal & Severe	Minor
WEEKEND	0 - Weekday	68,21 %	401 (6.65 %)	3712 (61.56 %)
	1- Weekend	31,79 %	197 (3.27 %)	1720 (28.52 %)
Time of day	MORNING	46,25 %	312 (5.17 %)	2477 (41.08 %)
	AFTERNOON	35,62 %	197 (3.27 %)	1951 (32.35 %)
	Otherwise (Night)	18,13 %	89 (1.48 %)	1004 (16.65 %)
Crash type	OBSTACLE COLLISION	8.61 %	68 (1.13 %)	451 (7.48 %)
	FALL	58.92 %	348 (5.77 %)	3205 (53.15 %)
	OVERTURNING	19,52 %	82 (1.36 %)	1095 (18.16 %)
	Otherwise	12,95 %	100 (1.66 %)	681 (11.29 %)

2.1.2. Rider information

The socio-demographic characteristics of riders involved in crashes provide valuable insights for urban planners and healthcare systems, facilitating the identification of high-risk groups and the formulation of relevant policies [6]. Therefore, 14 variables containing specific rider information were defined. These variables encompass information regarding the rider's gender, age, helmet usage, trip purpose, pre-crash manoeuvre, as well as the identified rider state and any potential traffic offences (Table 2).

The variable *GENDER* assumes the value 0 when the rider is female and 1 when it is male, as illustrated in Table 2.

Regarding the age of the rider, the following variables have been considered: (i) YOUNG18, when the rider is under 18 years old; (ii) ADULT65, when the rider is between 18 and 65 years old; and (iii) OLDER, when the rider is over 65 years old. When all three of these variables are simultaneously 0, then the age of the rider is unknown.

Table 2
Rider information variables.

Variables		Total (%)	Micromobility crashes	
			Fatal & Severe	Minor
Gender	0 - Female	23,20 %	104 (1.72 %)	1295 (21.48 %)
	1- Male	76,80 %	494 (8.19 %)	4137 (68.61 %)
Rider age	YOUNG18	9,24 %	58 (0.96 %)	499 (8.28 %)
	ADULT65	79,40 %	464 (7.69 %)	4324 (71.71 %)
	OLDER	6,78 %	63 (1.04 %)	346 (5.74 %)
	Otherwise (Unknown)	4,58 %	13 (0.22 %)	263 (4.36 %)
Helmet use	HELMET	31,19 %	219 (3.63 %)	1662 (27.56 %)
Trip purpose	LEISURE	25,89 %	196 (3.25 %)	1365 (22.64 %)
	COMMUTE	6,33 %	38 (0.63 %)	344 (5.70 %)
	PROFESSIONAL	0,30 %	1 (0.02 %)	17 (0.28 %)
	Otherwise	67,48 %	363 (6.02 %)	3706 (61.46 %)
Rider maneuver	EVASIVE MANEUVER	1,00 %	7 (0.12 %)	53 (0.88 %)
	BRAKING	1,49 %	3 (0.05 %)	87 (1.44 %)
	Otherwise	97,51 %	588 (9.75 %)	5292 (87.76 %)
Rider state and offences	ALCOHOL	0,90 %	3 (0.05 %)	51 (0.85 %)
	SPEED	3,67 %	38 (0.63 %)	183 (3.03 %)
	DISTRACTION	4,43 %	35 (0.58 %)	232 (3.85 %)
	SUDDEN ILLNESS	0,43 %	8 (0.13 %)	18 (0.30 %)

The variables *LEISURE*, *COMMUTE*, and *PROFESSIONAL* have been derived from the “Trip purpose” variable in the original database. When all three newly defined variables are equal to 0, it indicates that either the reason for travel is unknown (in 93.83 % of cases), or the travel is for another unspecified reason different from the ones mentioned (in 6.17 % of cases).

Additionally, the variables *EVASIVE MANEUVER* and BRAKING have been defined based on the “Rider maneuver” variable in the original database. When both new variables have a value of 0, it means that either the rider is following a straight path (50.44 %), or they are performing a different maneuver from the ones mentioned (11.02 %), or the specific type of maneuver being executed is unknown (38.54 %).

2.1.3. Vehicle information

The Vehicle is one of the most critical concurrent factors typically examined in road safety assessments. Consequently, two related variables were included in the analysis: *BICYCLE*, indicating whether the rider was on a bicycle or a PMD, and *VH_BADMAINTENANCE*, which takes on a value of 1 when there is substantial evidence of the vehicle being in poor condition (Table 3).

The variable *BICYCLE* was formulated using several vehicle-type data available from the DGT database. Micromobility vehicles were categorized as follows: (i) bicycle; (ii) e-scooters; (iii) other PMDs, including segways, skateboards, and electric wheelchairs; and (iv) unspecified PMD, covering personal mobility devices where data did not specify the type according to the available data. An analysis of crashes within category (iii) revealed that this category comprises only 0,065 % of the PMDs involved in crashes. Additionally, considering that e-scooters have been the predominant PMD in Spain in recent years, it is reasonable to assume that category (iv) primarily comprises e-scooters. Consequently, only two categories have been considered in this study: (i) bicycle (value 1) and (ii) PMDs (value 0), including the previous categories (ii), (iii), and (iv).

Incorporating vehicle information is crucial because, despite bicycles and other PMDs (mainly e-scooters) being classified as micromobility vehicles, disparities in kinetic response and damage mechanisms may exist between these vehicles. Therefore, it is essential to assess whether the type of micromobility vehicle is a determinant of injury severity for users of these vehicles.

2.1.4. Infrastructure information

The analysis of crash locations plays a vital role in enhancing safety planning and engineering solutions. In single micromobility crashes, the role of infrastructure could be relevant. Therefore, with the aim of identifying whether infrastructure-related variables could influence the severity of micromobility users, these variables have been included in the analysis, as shown in Table 4.

Nine infrastructure-related variables have been defined from the DGT database information, and they have been categorized into three categories: (i) crash location; (ii) rider location; and (iii) pavement condition.

Crash location has been further categorized into three groups: (i) crashes occurring at intersections (*INTERSECTION*); (ii) crashes

Table 3
Vehicle information variables.

Variables		Total (%)	Micromobility crashes	
			Fatal & Severe	Minor
Micromobility vehicle	0 - PMD	9.09 %	70 (1.16 %)	478 (7.93 %)
	1- Bicycle	90.91 %	528 (8.76 %)	4954 (82.16 %)
Vehicle bad maintenance	VH_BADMAINTENANCE	2.75 %	20 (0.33 %)	146 (2.42 %)

Table 4
Infrastructure information variables.

Variables		Total (%)	Micromobility crashes	
			Fatal & Severe	Minor
Crash location	INTERSECTION	13.43 %	92 (1.53 %)	718 (11.91 %)
	ROUNDABOUT	3.90 %	23 (0.38 %)	212 (3.52 %)
	Otherwise	82.67 %	483 (8.01 %)	4502 (74.66 %)
Rider location	VH_LANE	27.88 %	189 (3.13 %)	1492 (24.74 %)
	SHOULDER	0.25 %	2 (0.03 %)	13 (0.22 %)
	SIDEWALK	2.35 %	12 (0.20 %)	130 (2.16 %)
	BIKE SIDEWALK	2.34 %	18 (0.30 %)	123 (2.04 %)
	BIKE LANE	8.36 %	52 (0.86 %)	452 (7.50 %)
	BUS LANE	0.28 %	2 (0.03 %)	15 (0.25 %)
	Otherwise	58.54 %	323 (5.36 %)	3207 (53.18 %)
Bad pavement condition	BAD_PAVEMENT	17.73 %	129 (2.14 %)	940 (15.59 %)

occurring at roundabouts (*ROUNDABOUT*); and (iii) crashes occurring at midblock (variables *INTERSECTION* and *ROUNDABOUT* with a value of 0).

Rider location, referring to the infrastructure where the micromobility user was riding before the crash, has been defined with six categories: (i) *VH*_LANE, (ii) *SHOULDER*, (iii) *SIDEWALK*, (iv) *BIKE SIDEWALK*; (v) *BIKE_LANE*, and (vi) *BUS_LANE*. When all these variables have a value of 0, it signifies that either the rider was traveling in a different location from those mentioned, such as the tram lane or the median of the roadway (6.18 % of cases), or the database does not specify the location where the rider was traveling (93.82 %).

Finally, the variable BAD_PAVEMENT has been defined from the "pavement conditions" variable in the original database. The new variable takes the value 1 when the pavement is not dry or clean (e.g., the pavement is muddy, wet, icy, snowy, etc.), or when it has issues such as potholes. It takes the value 0 otherwise.

2.2. Random Forest

The mentioned variables were utilized in calibrating and validating a classification model employing the Random Forest methodology, with *SEVERITY* as the dependent variable. The selection of the Random Forest approach is justified by its robust classification capabilities, particularly its superior performance in the context of road crash injury severity prediction [26].

As previously noted, an issue arises due to the imbalance in the *SEVERITY* variable, where the number of crashes resulting in fatalities and severe injuries (598) is significantly lower than those causing minor injuries (5,432). Such data imbalance can potentially introduce bias into classifiers since they tend to the majority class [31]. To mitigate the risk of bias and address the classification accuracy challenges in the Random Forest model, it is necessary to tackle the data imbalance issue [32]. This was achieved by applying the Synthetic Minority Oversampling Technique (SMOTE).

The SMOTE technique stands out as one of the most influential and well-known data sampling and balancing algorithms in Machine Learning. It is an oversampling method specially designed to address imbalanced data [33]. This technique is based on creating new minority instances by generating random synthetic examples (new data). These newly generated instances result from interpolating between neighbouring minority class instances while preserving the original features [31–33]. To optimize model performance and mitigate any potential oversampling-related issues, critical hyperparameters were adjusted. These included parameters governing the oversampling rate of the minority class and the parameter dictating the number of nearest neighbours to consider during synthetic sample generation. A cross-validation process was conducted, involving multiple SMOTE executions employing different hyperparameter configurations. The objective was to comprehensively evaluate their impact on model performance and select the set of hyperparameters that yielded the best results. Once data had been successfully balanced through SMOTE, the Random Forest methodology was applied.

Random Forest is one of the most well-known and powerful Machine Learning supervised learning techniques for addressing classification or regression problems and making predictions [34]. It is typically applied to assess the discriminate power of all predictor variables in the model within the classification tree framework [35]. The model is constructed from a set of individual trees, each being trained on a bootstrap sample of the original data. Moreover, for each partition of the tree nodes, only one set of randomly selected independent variables is evaluated. Thus, the sample data is split into two child nodes to maximize the explained variance of the dependent variable [36–38]. Therefore, Random Forest involves a first random sample of the data and a second random sample of the predictor variables in each partition. This model-building process based on a set of independent trees helps to alleviate the overfitting issues identified in Classification and Regression Tree (CART) models, thereby improving classification accuracy [34,37,38].

The data that has not been included in the construction of each tree when selecting a bootstrap sample from the original data is termed Out of Bag (OOB) sample, and Random Forest makes predictions with these data to determine the accuracy of the model, calculating the OOB error [37].

Random Forest provides two different importance measures: mean decrease Gini (GINI) and Mean Decrease Accuracy (MDA), which can be used for ranking variables and for variable selection. GINI is the sum of all decreases in Gini impurity due to a given variable used to form a split in the Random Forest, normalized by the number of trees. MDA quantifies the importance of a variable based on the change in prediction accuracy (OOB error) when the values of the variable are randomly permuted compared to the original observations [39].

3. Results and discussion

3.1. Random Forest model

3.1.1. Model construction and validation

Ramdom Forest is the methodology used to construct a model for analysing single micromobility crashes. This model explores the relationship between crash outcome (*SEVERITY*) and several attributes related to crashes, riders, vehicles, and infrastructure (independent variables).

To build and subsequently validate the model, cross validation has been conducted. Two distinct datasets have been created from the original data: a first training set, consisting of 70 % of the original data selected randomly, and a second validation set, consisting of the remaining 30 % of the data. The first dataset was used to adjust the model parameters and construct it, while validation was carried out using the second dataset. The 70:30 training-to-set data ratio has been shown to yield superior performance scores across various tree-based machine learning models and has become the most frequently recommended and widely adopted option in the field [40].

To construct the Random Forest model and generate the resulting variable importance plots, it is essential to adjust the key parameters of the model. The number of variables used for node splitting (*mtry*) is, by default and for classification problems, $\sqrt{x}$, where x represents the number of variables. Additionally, the default number of trees (*ntree*) is 500. To optimize these model hyperparameters, Random Forest models have been built with varying values of *mtry* (ranging from $\sqrt{x}$ to 31) and *ntree* (from 50 to 500). The hyperparameters values that delivered the best model performance were determined to be *mtry=15* and *ntree=500.*

Fig. 4 illustrates the progression of the model OOB error as the value of *ntree* changes, along with the evolution of classification errors for fatal or seriously injured cases and for slightly injured ones. The OOB error curve, as well the classification error curves, demonstrate remarkable stability when *ntree=500*. The classification errors are more pronounced for slightly injured cases (25 %) than for fatal or seriously injured cases (10 %), with the model OOB error calculated at 18.04 %.

The performance of the model was evaluated using the confusion matrix that compares the fitted results with the actual observations [41], as depicted in Fig. 5. The diagonal elements of this matrix represent the count of correctly classified crashes for each category of injury severity, while the off-diagonal elements represent the count of misclassified crashes.

These results have been used to calculate the F-measure value, which approaches 1 when there is perfect precision and recall. The F-measure score is computed using the harmonic mean of precision and recall [42,43]. The following equations are employed to calculate precision, recall and F-measures for both fatal and seriously injured and slightly injured categories.

$$Precision_1 = \frac{True\ positive}{True\ positive + False\ positive} = \frac{1,379}{1,379 + 378} = 0.7849 \tag{1}$$

$$Recall_1 = \frac{True\ positive}{True\ positive + False\ negative} = \frac{1,379}{1,379 + 232} = 0.856 \tag{2}$$

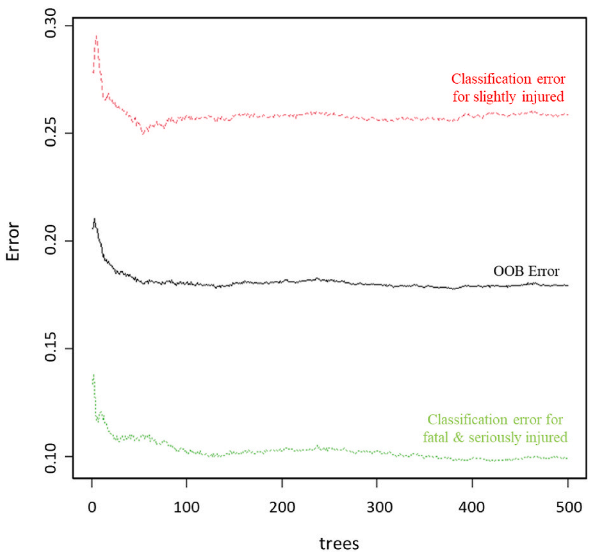

Fig. 4. OOB and classification Errors.

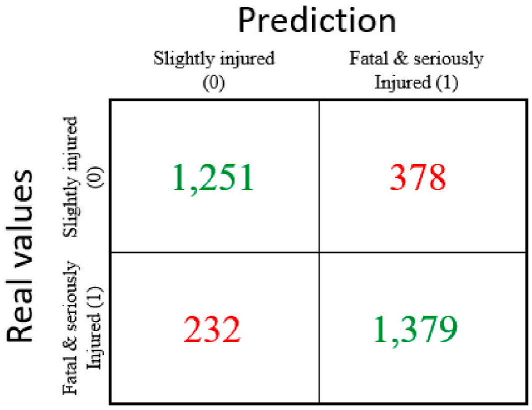

Fig. 5. Confusion matrix for Random Forest model using the validation set.

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} = 2 \cdot \frac{0.7849 \cdot 0.856}{0.7849 + 0.856} = \mathbf{0.8189} \tag{3}$$

$$Precision_2 = \frac{True\ positive}{True\ positive + False\ positive} = \frac{1,251}{1,251 + 232} = 0.8435 \tag{4}$$

$$Recall_2 = \frac{True\ positive}{True\ positive + False\ negative} = \frac{1,251}{1,251 + 378} = 0.768 \tag{5}$$

$$F_2 = 2 \cdot \frac{precision \cdot recall}{precision + recall} = 2 \cdot \frac{0.8435 \cdot 0.768}{0.8435 + 0.768} = \mathbf{0.8039} \tag{6}$$

The calculated F-measures are 0.8189 for fatal and seriously injured and 0.8039 for slightly injured. These values, which are very close to 1, also confirm the accuracy of the model.

Furthermore, Fig. 6 shows the ROC curve for predicting the validation data using the Random Forest model (model classifier). For comparison, it also includes the ROC curves of the perfect classifier and the classifier without predictive value. The closer the model classifier is to the perfect classifier, the better the model performs. Additionally, AUC (Area Under the Curve) is a valuable metric for assessing the quality of class separation in classifiers [27,44]. The model performs well when the AUC is close to one. The AUC for the developed Random Forest model is 0.875, indicating a strong performance of the classification model.

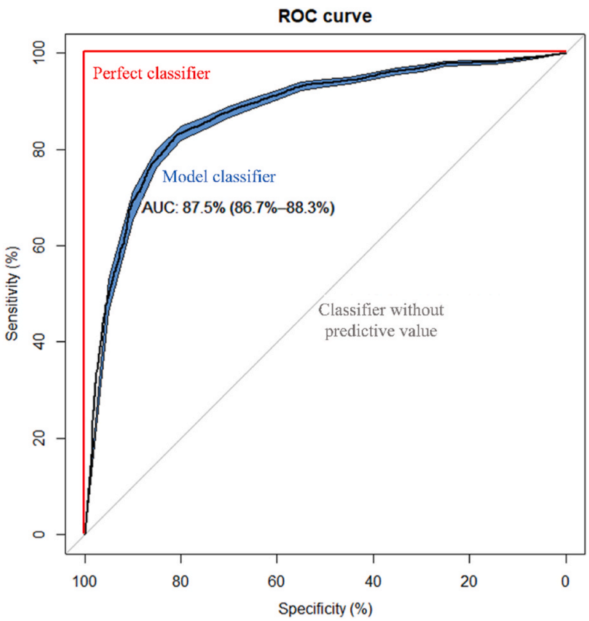

Fig. 6. ROC curve for RF model.

The results confirm the performance and validation of the model while also providing information about its uncertainty.

The developed and validated Random Forest model provides two variable importance plots, based on the mean decrease Gini (GINI) and Mean Decrease Accuracy (MDA) criteria. These rankings elucidate the importance of each variable in the classification model. The results from both rankings have been analyzed and compared. Notably, there is a high correlation (0.88) between the importance rankings established by both criteria, yielding similar results, as expected. Additionality, ten Random Forest models (with the same hyperparameters) were created to assess the robustness and stability of the model along with the variable importance criteria. The GINI index was found to provide more consistent results than the MDA, which aligns with findings from other studies [39]. Consequently, the variable importance ranking from Random Forest as determined by the GINI index, has been used (Fig. 7).

In Random Forest, the GINI index provides insights into which features are the most crucial for accurately classifying data and, as a result, are more influential in the model decision-making.

Based on the GINI index results, the primary factors influencing crash classification by injury severity, include the day of the week and whether the rider is wearing a helmet. Also significant are whether the rider is using motor vehicle lanes, riding at an intersection, the pavement condition, the timing of the crash (afternoon or not), and the purpose of the trip (leisure or non-leisure).

Furthermore, although to a lesser extent, other factors contributing to data classification include whether the micromobility vehicle is in a bike lane, whether the crash resulted from a fall, the type of vehicle involved (bicycle or PMD, mainly e-scooters), the presence of collisions with obstacles, and certain rider-related attributes such as gender, age, and specific offences, particularly distractions and speeding.

Other factors, such as alcohol consumption and whether the rider was on the sidewalk or in a bus lane, hold less significance in the classification process.

3.1.2. Decision rules

In the previous section, the most and least relevant variables for classifying crashes based on injury severity were individually identified. However, crashes, along with their consequences, typically result from the interaction of multiple factors. Therefore, it is crucial to determine the likely injury severity of a crash when several factors are concurrently present. This information is vital for authorities to make informed decisions regarding micromobility road safety and allocate available resources effectively to reduce the crash rate. Decision rules can be valuable in this regard, as emphasized by other researchers [28]. Out of the 62 decision rules generated by the Random Forest model, Tabla 5 displays the most significant rules that can assist planning, design, and management authorities in their decision-making process.

After a comprehensive analysis of all decision rules generated by the model, the following conclusions can be drawn.

- The likelihood of a crash being fatal or resulting in serious injuries increases when falls occur on a poorly conditioned pavement. All crashes with both of these factors have been classified as fatal or serious by the model (decision rule 10 and other rules not included in Table 5). If only one of these factors is present, the crash severity depends on the interaction with other factors (decision rule 19 and other rules not listed in Table 5). Poor pavement conditions have previously been identified as a significant cause of e-cyclist crashes [20], but our study extends this finding to micromobility crashes in general, particularly those involving falls.

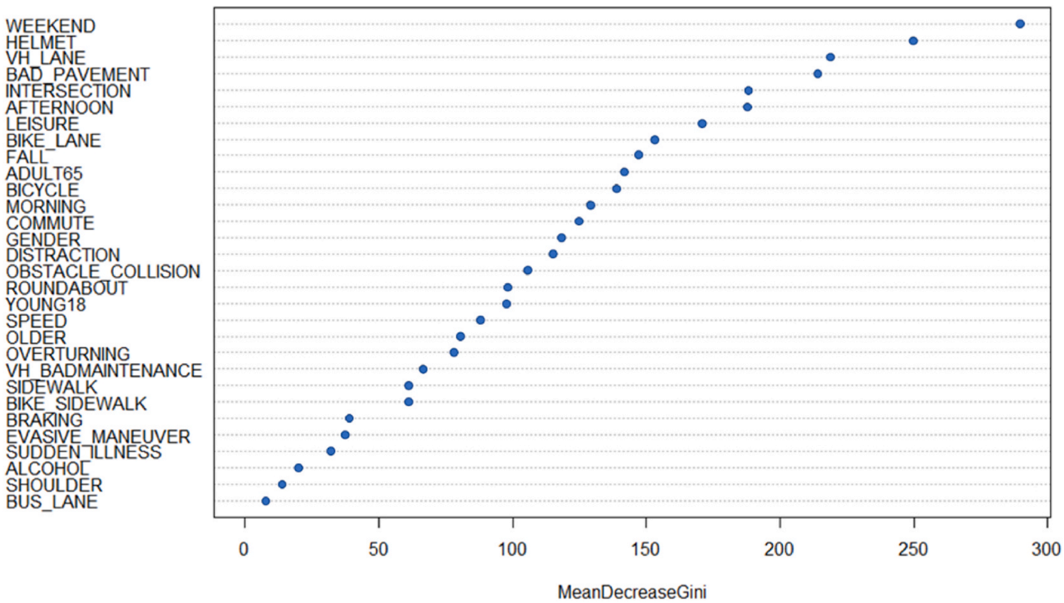

Fig. 7. Variable importance plot (GINI).

Table 5
The main decision rules identified by the RF model.

Number	Decision rules	THEN	Frequency (%)	Error (%)
1	IF (WEEKEND = 0) AND (MORNING = 1) AND (VH_LANE = 0) AND (BIKE_LANE = 0) AND (LEISURE = 0) AND (SPEED = 1)	Fatal or seriously injured	1.04	5.35
2	IF (WEEKEND = 1) AND (MORNING = 0) AND (BICYCLE = 1) AND (BIKE_LANE = 0) AND (BIKE_SIDEWALK = 0) AND (LEISURE = 1)	Slightly injured	1.06	11.40
3	IF (ROUNDABOUT = 0) AND (BICYCLE = 0) AND (OVERTURNING = 0) AND (SIDEWALK = 0) AND (PROFESSIONAL = 0) AND (SPEED = 1)	Fatal or seriously injured	1.05	14.16
4	IF (OBSTACLE_COLLISION = 0) AND (FALL = 0) AND (OVERTURNING = 0) AND (BIKE_LANE = 0) AND (HELMET = 1) AND (LEISURE = 0)	Fatal or seriously injured	2.64	16.14
5	IF (MORNING = 1) AND (INTERSECTION = 1) AND (EVASIVE_MANEUVER = 0) AND (VH_LANE = 1) AND (GENDER = 1) AND (LEISURE = 1)	Fatal or seriously injured	1.22	16.67
6	IF (BICYCLE = 0) AND (OBSTACLE_COLLISION = 0) AND (OVERTURNING = 0) AND (GENDER = 0) AND (HELMET = 0) AND (LEISURE = 1)	Fatal or seriously injured	1.17	16.67
7	IF (WEEKEND = 1) AND (BICYCLE = 1) AND (VH_LANE = 1) AND (GENDER = 1) AND (LEISURE = 0) AND (SPEED = 0)	Slightly injured	1.69	16.94
8	IF (AFTERNOON = 1) AND (BICYCLE = 1) AND (ADULT65 = 1) AND (GENDER = 0) AND (HELMET = 0) AND (LEISURE = 0)	Slightly injured	2.58	18.99
9	IF (MORNING = 1) AND (OVERTURNING = 1) AND (BIKE_SIDEWALK = 0) AND (OLDER = 0) AND (GENDER = 0) AND (SPEED = 0)	Slightly injured	1.33	22.22
10	IF (MORNING = 1) AND (FALL = 1) AND (BAD_PAVEMENT = 1) AND (VH_LANE = 0) AND (GENDER = 1) AND (LEISURE = 1)	Fatal or seriously injured	1.61	22.99
11	IF (FALL = 1) AND (OVERTURNING = 0) AND (VH_LANE = 0) AND (GENDER = 1) AND (LEISURE = 0) AND (COMMUTE = 1)	Fatal or seriously injured	1.47	23.27
12	IF (MORNING = 1) AND (INTERSECTION = 0) AND (VH_LANE = 1) AND (OLDER = 1) AND (COMMUTE = 0) & (DISTRACTION = 0)	Fatal or seriously injured	1.95	24.64
13	IF (BICYCLE = 1) AND (BAD_PAVEMENT = 0) AND (ADULT65 = 1) AND (GENDER = 0) AND (HELMET = 0) AND (LEISURE = 1)	Fatal or seriously injured	1.25	25.19
14	IF (MORNING = 0) AND (AFTERNOON = 0) AND (BICYCLE = 0) AND (FALL = 0) AND (GENDER = 1) AND (SPEED = 0)	Fatal or seriously injured	1.04	26.79
15	IF (INTERSECTION = 0) AND (BICYCLE = 0) AND (GENDER = 0) AND (LEISURE = 0) AND (SPEED = 0) AND (DISTRACTION = 0)	Slightly injured	1.28	28.99
16	IF (ROUNDABOUT = 0) AND (BICYCLE = 1) AND (BIKE_SIDEWALK = 0) AND (YOUNG18 = 0) AND (HELMET = 1) AND (SPEED = 1)	Fatal or seriously injured	1.17	29.37
17	IF (MORNING = 0) AND (AFTERNOON = 0) AND (OVERTURNING = 0) AND (OLDER = 0) AND (HELMET = 1) AND (SPEED = 0)	Slightly injured	1.24	28.36
18	IF (WEEKEND = 1) AND (MORNING = 1) AND (ADULT65 = 0) AND (GENDER = 1) AND (LEISURE = 0) AND (SPEED = 0)	Fatal or seriously injured	1.16	28.80
19	IF (BICYCLE = 1) AND (FALL = 1) AND (OVERTURNING = 0) AND (YOUNG18 = 0) AND (GENDER = 0) AND (COMMUTE = 0)	Slightly injured	1.44	29.49
20	IF (MORNING = 1) AND (INTERSECTION = 0) AND (FALL = 1) AND (OVERTURNING = 0) AND (GENDER = 0) AND (LEISURE = 0)	Slightly injured	2.72	31.63
21	IF (WEEKEND = 0) AND (FALL = 1) AND (OVERTURNING = 0) AND (BAD_PAVEMENT = 0) AND (GENDER = 1) AND (HELMET = 1)	Slightly injured	3.36	37.74
22	IF (MORNING = 1) AND (VH_BAD_MAINTENANCE = 1) AND (VH_LANE = 1) AND (GENDER = 1) AND (LEISURE = 1) & (SPEED = 0)	Fatal or seriously injured	3.23	38.11

- Micromobility crashes analyzed involving elderly riders (over 65 years) tend to be more serious than those not involving these riders (decision rules 9, 12, 17 and other rules not included in Table 5). This aligns with findings from other studies on single-vehicle crashes involving bicycles and e-bikes [16,20].
- Micromobility crashes occurring on weekdays in the morning, outside of vehicle and bike lanes, and involving riders not travelled for leisure, but committing a speed offence, are much more likely to be classified as serious or fatal (decision rule 1).
- An adult woman (aged 18 to 65) involved in a bicycle crash while riding for leisure without a helmet is more likely to be classified as seriously injured or fatal. However, if this woman rides a bicycle in the afternoon, without a helmet, and for a non-leisure purpose, her crash is more likely to be classified as slightly injured (decision rules 8 and 13). In general, traveling for leisure seems to increase the injury severity of crashes, especially when the rider is not wearing a helmet [16].
- Micromobility crashes involving only one PMD (mainly e-scooter) appear to be more serious than those involving only one bicycle, especially at night (decision rules 3, 6, 14 and other rules not included in Table 5).
- Falls involving males wearing helmets on weekdays are more likely to result in minor injuries. In general, helmet use has been observed to decrease the severity of crashes, as expected (decision rule 21 and other rules not listed in Table 5).
- The likelihood of a crash being fatal or resulting in serious injuries increases for men involved in falls while commuting and riding outside of vehicle lanes (decision rule 11). This finding consistent with another study that identified using e-bike primarily for commuting to work or school as a risk factor in e-bicycle crashes, although it did not specify the infrastructure where the cyclist was riding [20].

- Bicycle crashes occurring on weekends, outside of the morning hours, and for leisure purposes are more likely to be classified as minor (decision rule 2).
- The likelihood of a crash being fatal or resulting in serious injuries is higher in collisions with animals, run-off crashes, and falling off cliffs, even if the rider is wearing a helmet (decision rule 4 and other rules not included in Table 5).
- In general, it has been observed that speed offences increase the severity of crashes (decision rules 1, 3, 16 and other rules not included in Table 5). This aligns with the findings of another study focused on e-cyclist [20].
- The likelihood of a crash being fatal or resulting in serious injuries is higher when it occurs at intersections compared to roundabouts or other locations. The severity depends on the simultaneous occurrence of other factors (decision rule 5 and other rules not listed in Table 5).

This study represents a significant contribution to the road safety field due to the extensive sample of analyzed crashes. It includes data from crashes reported by the police, encompassing all injury-related crashes that occurred in Spain. The use of this database, as opposed to relying on surveys [20] or hospital records, which may potentially underestimate data [9], enhances the accuracy of the results.

Furthermore, this research encompasses all micromobility vehicles (including bicycles and PMDs) involved in single micromobility crashes. Additionally, it takes into account a comprehensive set of factors related to riders, micromobility vehicles, and the crashes themselves. In terms of crashes characteristics, the study not only considers the crash type and when it occurred, but also where it occurred. Some studies have suggested a potential link between e-scooter crashes severity and infrastructure [6], yet comprehensive studies in this regard have been limited [45]. Consequently, this study delves into infrastructure-related factors as well.

However, it is important to acknowledge the study limitations, which could be addressed in future research. The database used does not include crashes that result in only property damage. Therefore, single micromobility crashes might be underrepresented, like what occurs with SBCs [13]. Additionally, despite the substantial growth in PMD demand, particularly e-scooters, in recent years, further expansion is expected. Consequently, a similar study should be carried out in a few years to analyze how these factors evolve in their influence on single micromobility crashes injuries. Furthermore, certain variables lack available information and have consequently not been considered. For example, detailed information regarding road design or traffic density is unavailable. The latter variable could be particularly relevant in crashes that occur within vehicle lanes. Therefore, future analyses should explore the impact of these variables as well.

4. Conclusions

The substantial increase in micromobility modes of transportation has changed mobility patterns and urban lifestyles worldwide. However, this surge in micromobility users has also led to a rise in crashes involving these vehicles. In 2020, 26.51 % of crashes involving micromobility users occurred without any interaction with other users (single micromobility crashes). This proportion has steadily increased from 16.91 % in 2016 to 26.51 % in 2020. These single micromobility crashes primarily fall into four categories: (i) fall; (ii) overturning; (iii) obstacle collision; and (iv) other, which includes collisions with animals, run-off crashes, and other crashes. Around 10 % of these crashes result in serious injuries or fatalities. Therefore, it is worthy to study the factors influencing these collisions and their impact on the crash severity.

This study focused on crashes that occurred in Spanish urban areas from 2016 to 2020, using data from the National Registry of Traffic Accidents Victims. Specifically, it analyzes crashes involving only one micromobility user (cyclist, e-scooter rider, or other PMD rider) without any interaction with other users.

To evaluate the influence of several factors – such as crash characteristics, socio-demographic characteristics and behavior of riders, type and maintenance of the vehicle, and infrastructure conditions –on crash severity, a model based on Random Forest has been developed, calibrated, and validated. Prior to model calibration, data were balanced using the SMOTE technique to mitigate biases stemming from differences in the sample sizes of slightly injured crashes versus serious and fatal crashes.

According to the GINI index, among the ten most influential variables in classifying the severity of single micromobility crashes, three are related to the collision (*WEEKEND*, *AFTERNOON*, *FALL*), three to the rider (*HELMET*, *LEISURE*, *ADULT65*), and four to infrastructure (*VH_LANE*, *BAD_PAVEMENT*, *INTERSECTION*, *BIKE_LANE*). The variable distinguishing between bicycles and other micromobility vehicles occupies the 11th position in terms of importance.

Moreover, decision rules have been obtained to study the combination of multiple factors.

Considering micromobility vehicles, this study makes a clear distinction between bicycles and PMDs, with stand-up e-scooters as a prominent type within the PMD category. Of all the decision rules uncovered which involved the type of micromobility vehicle, bicycles were featured in nearly 70 % of them, while PMDs constituted the remaining 30 %.

For single-bicycle crashes, the severity tends to increase, especially for adult women on leisure trips without helmets and for adults who wear helmets but engage in excessive speed. In contrast, single PMD crashes tend to be more severe when excessive speed is involved, the trip is for leisure, and it occurs at night. Helmet use and leisure trips significantly influence the injury severity for users of both micromobility vehicles. However, it has been observed that single-crashes occurring at night have a worse outcome when the micromobility user is on a PMD compared to a bicycle. Furthermore, generally, it can be concluded that the proportion of decision rules resulting in severe or fatal collisions is higher for single PMD crashes compared to single-bicycle crashes.

Without distinguishing between the type of micromobility vehicle, the decision rules reveal that, in general, collisions involving elderly riders (over 65 years of age) tend to result in fatal or severe injuries. In contrast, for younger people, the severity of the collision depends also on other factors. For instance, a rider aged 18 to 65 who has a single micromobility crash during a leisure trip is more likely to sustain severe injuries than if the crash occurs when riders are traveling for other purposes.

Regarding the gender, there is a considerable number of decision rules in which this variable appears, with a higher proportion of fatal or severe crashes for males than for females, particularly in the morning.

Regarding the four types of single micromobility crashes, the category that encompasses collision with animals, run-off crashes, and falling off cliffs tends to result in more severe outcomes even when the rider is wearing a helmet. Falls tend to be serious when the pavement is muddy, wet, icy, or snowy, or when there are potholes. In contrast, overturning crashes tend to have slightly less severe consequences. These consequences are exacerbated when crashes occur in motor vehicle lanes rather than in bike lanes or bike paths.

These findings can serve as a foundation for enhancing micromobility road safety, preventing single micromobility crashes, and mitigating their consequences. Authorities should prioritize dedicated infrastructure for micromobility users, separated from motorized vehicle traffic. These lanes should have well-maintained pavement and remain free of obstacles to minimize the severity of consequences in the event of a fall. Furthermore, campaigns promoting helmet use for these vehicles, discouraging distractions while riding, and especially enforcing speed limits should be intensified. By addressing both infrastructure and human factors in this manner, the occurrence of fatal or serious single micromobility crashes could be significantly decreased.

Data availability statement

The data employed in this study were supplied by the Dirección General de Tráfico of Spain. Regrettably, the authors do not have permission to share data.

CRediT authorship contribution statement

Almudena Sanjurjo-de-No: Writing - review & editing, Writing - original draft, Visualization, Validation, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Ana María Pérez-Zuriaga:** Writing - review & editing, Visualization, Validation, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Alfredo García:** Validation, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research is part of the research project PID2019-111744RB-I00, funded by MCIN/AEI/10.13039/501100011033.

Likewise, this research has been partially funded by the European Union-NextGenerationEU (RD 289/2021) through the "Margarita Salas" grant awarded to Almudena Sanjurjo (UP2021-035), a researcher at the Universidad Politécnica de Madrid (UPM) to support her stay at the Universitat Politècnica de València (UPV).

Finally, the authors would like to also thank to the Dirección General de Tráfico (DGT) for providing the databases.

References

[1] A. Santacreu, G. Yannis, O. de Saint Leon, P. Crist, Safe micromobility. International transport forum. https://www.itf-oecd.org/sites/default/files/docs/safe-micromobility_1.pdf, 2020. (Accessed 14 March 2023).

[2] J. Zagorskas, M. Burinskienė, Challenges caused by increased use of E-powered personal mobility vehicles in European cities, Sustain. Times 12 (1) (2019) 273, https://doi.org/10.3390/su12010273.

[3] C. Gioldasis, Z. Christoforou, R. Seidowsky, Risk-taking behaviors of e-scooter users: a survey in Paris, Accid. Anal. Prev. 163 (2021), 106427, https://doi.org/10.1016/j.aap.2021.106427.

[4] C. Hardt, K. Bogenberger, Usage of e-scooters in urban environments, Transport. Res. Procedia 37 (2019) 155–162, https://doi.org/10.1016/j.trpro.2018.12.178.

[5] C. Garman, S.G. Como, I.C. Campbell, J. Wishart, K. O'Brien, S. McLean, Micro-mobility vehicle dynamics and rider kinematics during electric scooter riding, in: Proceedings of the WCX SAE World Congress Experience 2020, 2020, pp. 21–23. Detroit, MI, USA.

[6] K. Kazemzadeh, M. Haghani, F. Sprei, Electric scooter safety: an integrative review of evidence from transport and medical research domains, Sustain. Cities Soc. 89 (2023), 104313, https://doi.org/10.1016/j.scs.2022.104313.

[7] J.B. Cicchino, P.E. Kulie, M.L. McCarthy, Severity of e-scooter rider injuries associated with trip characteristics, J. Saf. Res. 76 (2021) 256–261, https://doi.org/10.1016/j.jsr.2020.12.016.

[8] C. Jones, B. Hammig, E-scooter injuries reported to US emergency Rooms in 2020 epidemiology and injury mechanisms and helmet use, J. of Emerg. Med., Trauma & Surg. Care 3 (1) (2021) 1–3.

[9] T.K. Trivedi, C. Liu, A.L.M. Antonio, N. Wheaton, V. Kreger, A. Yap, D. Schriger, J.G. Elmore, Injuries associated withstanding electric scooter use, JAMA Netw. Open 2 (1) (2019), https://doi.org/10.1001/jamanetworkopen.2018.7381.

[10] European Transport Safety Council (Etsc), How safe is walking and cycling in Europe?. PIN Flash Report 38, Available in (14/03/2023): https://etsc.eu/wp-content/uploads/PIN-Flash-38_FINAL.pdf, 2020.

[11] P. Schepers, N. Agerholm, E. Amoros, R. Benington, T. Bjørnskau, S. Dhondt, B. Geus, C. Hagemeister, B.P.Y. Loo, A. Niska, An international review of the frequency of single-bicycle crashes (SBCs) and their relation to bicycle modal share, Inj. Prev. 21 (2015) 138–143, https://doi.org/10.1136/injuryprev-2013-040964.

[12] S. Boufous, L. de Rome, T. Senserrick, R.Q. Ivers, Single- versus multi-vehicle bicycle road crashes in Victoria, Australia, Inj. Prev. 19 (2013) 358–362, https://doi.org/10.1136/injuryprev-2012-040630.

[13] R. Utriainen, S. O'Hern, M. Pöllänen, Review on single-bicycle crashes in the recent scientific literature, Transp. Rev. (2022), https://doi.org/10.1080/01441647.2022.2055674.
[14] M.J. Boele-Vos, K. Van Duijvenvoorde, M.J.A. Doumen, C.W.A.E. Duivenvoorden, W.J.R. Louwerse, R.J. Davidse, Crashes involving cyclists aged 50 and over in The Netherlands: an in-depth study, Accid. Anal. Prev. 105 (2017) 4–10, https://doi.org/10.1016/j.aap.2016.07.016.
[15] B. Algurén, M. Rizzi, In-depth understanding of single bicycle crashes in Sweden - crash characteristics, injury types and health outcomes differentiated by gender and age-groups, J. of Transp. & Health 24 (2022), 101320, https://doi.org/10.1016/j.jth.2021.101320.
[16] J. Eriksson, A. Niska, A. Forsman, Injured cyclists with focus on single-bicycle crashes and differences in injury severity in Sweden, Accid. Anal. Prev. 165 (2022), 106510, https://doi.org/10.1016/j.aap.2021.106510.
[17] A.V. Olesen, T.K.O. Madsen, T. Hels, M. Hosseinpour, H.S. Lahrmann, Single-bicycle crashes: an in-depth analysis of self-reported crashes and estimation of attributable hospital cost, Accid. Anal. Prev. 161 (2021), 106353, https://doi.org/10.1016/j.aap.2021.106353.
[18] V. Martínez-Ruiz, P. Lardelli-Claret, E. Jiménez-Mejías, C. Amezcua-Prieto, J.J. Jiménez-Moleón, J.D. Luna del Castillo, Risk factors for causing road crashes involving cyclists: an application of a quasi-induced exposure method, Accid. Anal. Prev. 51 (2013) 228–237, https://doi.org/10.1016/j.aap.2012.11.023.
[19] J.P. Schepers, E. Fishman, P. den Hertog, K. Klein Wolt, A.L. Schwab, The safety of electrically assisted bicycles compared to classic bicycles, Accid. Anal. Prev. 73 (2014) 174–180, https://doi.org/10.1016/j.aap.2014.09.010.
[20] P. Hertach, A. Uhr, S. Niemann, M. Cavegn, Characteristics of single-vehicle crashes with e-bikes in Switzerland, Accid. Anal. Prev. 117 (2018) 232–238, https://doi.org/10.1016/j.aap.2018.04.021.
[21] S. Heydari, M. Forrest, J. Preston, Investigating the association between neighbourhood characteristics and e-scooter safety, Sustain. Cities Soc. 83 (2022), 103982, https://doi.org/10.1016/j.scs.2022.103982.
[22] S.N.F. Blomberg, O.C.M. Rosenkrantz, F. Lippert, H.C. Christensen, Injury from electric scooters in Copenhagen: a retrospective cohort study, BMJ Open 9 (2019) 1–8, https://doi.org/10.1136/bmjopen-2019-033988.
[23] H. Yang, Q. Ma, Z. Wang, Q. Cai, K. Xie, D. Yang, Safety of micro-mobility: analysis of E-Scooter crashes by mining news reports, Accid. Anal. Prev. 143 (2020), 105608, https://doi.org/10.1016/j.aap.2020.105608.
[24] M. Aizpuru, K.X. Farley, J.C. Rojas, R.S. Crawford, T.J. Moore, E.R. Wagner, Motorized scooter injuries in the era of scooter-shares: a review of the national electronic surveillance system, Am. J. Emerg. Med. 37 (6) (2019) 1133–1138, https://doi.org/10.1016/j.ajem.2019.03.049.
[25] A. Badeau, C. Carman, M. Newman, J. Steenblik, M. Carlson, T. Madsen, Emergency department visits for electric scooter-related injuries after introduction of an urban rental program, Am. J. Emerg. Med. 37 (8) (2019) 1531–1533, https://doi.org/10.1016/j.ajem.2019.05.003.
[26] K. Santos, J.P. Dias, C. Amado, A literature review of machine learning algorithms for crash injury severity prediction, J. Saf. Res. 80 (2022) 254–269, https://doi.org/10.1016/j.jsr.2021.12.007.
[27] A. Azhar, N.M. Ariff, M.A.A. Bakar, A. Roslan, Classification of driver injury severity for Accidents involving heavy vehicles with decision tree and random forest, Sustain. Times 14 (2022) 4101, https://doi.org/10.3390/su14074101.
[28] M. Ijaz, L. Ian, M. Zahid, A. Jamal, A comparative study of machine learning classifiers for injury severity prediction of crashes involving three-wheeled motorized rickshaw, Accid. Anal. Prev. 154 (2021), 106094, https://doi.org/10.1016/j.aap.2021.106094.
[29] G.D. Booth, M.J. Niccolucci, E.G. Schuster, Identifying Proxy Sets in Multiple Linear Regression: an Aid to Better Coefficient Interpretation. Research Paper INT (USA), U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Ogden, UT, USA, 1994.
[30] M. Sandri, P. Zuccolotto, A bias correction algorithm for the Gini variable importance measure in classification trees, J. Comput. Graph Stat. 17 (3) (2008) 611–628, https://doi.org/10.1198/106186008X344522.
[31] P. Skryjomski, B. Krawczyk, Influence of minority class instance types on SMOTE imbalanced data oversampling, Proceedings of the First International Workshop on Learning with Imbalanced Domains: Theory and Applications 74 (2017) 7–21. Available from: (14/03/2023): https://proceedings.mlr.press/v74/skryjomski17a.html.
[32] A. Fernández, S. García, F. Herrera, SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary, J. Artif. Intell. Res. 61 (2018) 863–905, https://doi.org/10.1613/jair.1.11192.
[33] A. Ishaq, S. Sadiq, M. Umer, S. Ullah, S. Mirjalili, V. Rupapara, M. Nappi, Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques, IEEE Access 9 (2021) 39707–39716, https://doi.org/10.1109/ACCESS.2021.3064084.
[34] Y. Liu, Y. Wang, J. Zhang, New machine learning algorithm: random forest, in: B. Liu, M. Ma, J. Chang (Eds.), Information Computing and Applications. ICICA 2012. Lecture Notes in Comput. Sci. 7473, Springer, Berlin, Heidelberg, 2012, https://doi.org/10.1007/978-3-642-34062-8_32.
[35] N.A. Dell, M.G. Vaughn, S.P. Srivastava, A. Alsolami, C.P. Salas-Wright, Correlates of cannabis use disorder in the United States: a comparison of logistic regression, classification trees, and random forests, J. Psychiatr. Res. 151 (2022) 590–597, https://doi.org/10.1016/j.jpsychires.2022.05.021.
[36] C. Nguyen, Y. Wang, H.N. Nguyen, Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic, J. Biomed. Sci. Eng. 6 (5) (2013) 551–560, https://doi.org/10.4236/jbise.2013.65070.
[37] M. Schonlau, R.Y. Zou, The random forest algorithm for statistical learning, STATA J.: Promot. Commun. on Stat. and Stata 20 (1) (2020) 3–29, https://doi.org/10.1177/1536867X20909688.
[38] A.M. Walker, A. Cliff, J. Romero, M. Shah, P. Jones, J.G.F. Machado Gazolla, D. Jacobson, D. Kainer, Evaluating the performance of random forest and iterative random forest based methods when applied to gene expression data, Comput. Struct. Biotechnol. J. 20 (2022) 3373–3386, https://doi.org/10.1016/j.csbj.2022.06.037.
[39] M.L. Calle, V. Urrea, Letter to the editor: stability of random forest importance measures, Briefings in Bioinform 12 (1) (2011) 86–89, https://doi.org/10.1093/bib/bbq011.
[40] S. Demir, E.K. Sahin, Comparison of tree-based machine learning algorithms for predicting liquefaction potential using canonical correlation forest, rotation forest, and random forest based on CPT data, Soil Dyn. Earthquake Eng. 154 (2022), 107130, https://doi.org/10.1016/j.soildyn.2021.107130.
[41] L. Lam, D.T. Do, D. Diep, D. Nguyet, Q.D. Truong, T.T. Tri, H.N. Thanh, N.Q.K. Le, Molecular subtype classification of low-grade gliomas using magnetic resonance imaging-based radiomics and machine learning, NMR Biomed. 35 (11) (2022), e4792, https://doi.org/10.1002/nbm.4792.
[42] Q.H. Kha, Q.T. Ho, N.Q.K. Le, Identifying SNARE proteins using an alignment-free method based on multiscan convolutional neural network and PSSM profiles, J. Chem. Inf. Model. 62 (2022) 4820–4826, https://doi.org/10.1021/acs.jcim.2c01034.
[43] C. Morris, J.J. Yang, Effectiveness of resampling methods in coping with imbalanced crash data: crash type analysis and predictive modeling, Accid. Anal. Prev. 159 (2021), 106240, https://doi.org/10.1016/j.aap.2021.106240.
[44] J.A. Cook, ROC curves and nonrandom data, Pattern Recognit. Lett. 85 (2017) 35–41, https://doi.org/10.1016/j.patrec.2016.11.015.
[45] N.R. Shah, S. Aryal, Y. Wen, C.R. Cherry, Comparison of motor vehicle-involved e-scooter and bicycle crashes using standardized crash typology, J. Saf. Res. 77 (2021) 217–228, https://doi.org/10.1016/j.jsr.2021.03.005.