Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 71 (2023) 6–13

Transportation Research Procedia

www.elsevier.com/locate/procedia

XV Conference on Transport Engineering, CIT2023

Analysis of the Elderly Pedestrian Injury Severity in Urban Traffic Accidents in Spain using Machine Learning Techniques

Daniel Gálvez-Pérez[a]*, Begoña Guirao[a], Armando Ortuño[b]

[a]*ETSI Caminos, Canales y Puertos, Universidad Politécnica de Madrid, Calle del Profesor Aranguren, 3, 28040 Madrid, Spain*
[b]*Escuela Politécnica Superior, Universidad de Alicante, San Vicente del Raspeig, 03690Alicante, Spain*

Abstract

Walking is an essential activity for a healthy lifestyle. In urbanized areas, the risk of suffering a traffic accident as a pedestrian is a matter of concern, and this situation has been widely studied. Elderly pedestrians are vulnerable road users due to their fragility and loss of physical faculties, and the probability of elderly pedestrians being killed or seriously injured in a traffic accident is higher than for the rest of the pedestrians. Furthermore, the elderly population is expected to increase in the coming decades. Hence, the study of these situations is necessary to build safer, healthier, and more sustainable cities. Little research has been devoted to the study of the relationship between injury severity elderly pedestrians and the characteristics of the accident location.

The objective of this study is to investigate the influence of accident and built environment factors on the severity of elderly pedestrian urban traffic accidents in Spain. For this purpose, logistic regression, and random forest models together with data resampling techniques were used to analyze the injury severity of vehicle collisions suffered by elderly pedestrians in Spain from 2016 to 2019, and its link with accident and built environment features. As expected, the random forest outperformed the logistic regression performance for both samples, but the logistic regression results are easier to interpret. Results showed the influence of both accident and built environment variables in the injury severity, being accidents in more populated cities less severe for all pedestrians. In addition, the age of the pedestrian, which indicates the need to study this age group in a more disaggregated manner, and accidents in dark spots artificially lighted, which suggests that those locations are not properly lighted for the elderly, are key factors for the injury level of elderly pedestrian traffic accidents.

Keywords: road safety; injury severity; elderly pedestrian; built environment; machine learning; imbalanced data

* Corresponding author.
E-mail address: daniel.galvezp@upm.es

2352-1465
10.1016/j.trpro.2023.11.051

1. Introduction

Vehicle-pedestrian collisions, like all traffic accidents, have been extensively studied from two complementary perspectives: occurrence, which assess the number of accidents registered within an area, or road element; and injury severity, which analyze the severity of the accident, normally in terms of the injury level of the victims.

The demographic ageing of the population has a significant effect on road safety outcomes, as supported by official statistical data. For instance, Europe has experienced a rise in road deaths among the elderly (above 65), from 17% of all road deaths in 1992 to 29% in 2018; and 50% of the dead pedestrians pertain to this age group (European Commission, 2021a). A high proportion of these fatalities were suffered by elderly pedestrians, with 34% in Europe and 44% in Spain. Regarding the location of the accident, elderly fatalities are more common in urban roads (53%) than all ages fatalities (38%). Elderly fatalities are relatively more frequent than all ages fatalities during the morning and afternoon, with two peaks between 9:00-11:00 and 15:00-18:00 (European Commission, 2021b). In Spain, with one of the oldest populations globally, elderly pedestrian fatalities on urban roads represented 70% of all pedestrian fatalities in 2019, while seniors accounted for only 19% of the population (DGT, 2020). This illustrates the inherent fragility of the elderly population. The ageing process normally causes the increase of impairments in population which raises their injury risk in case of a vehicle collision and reduces their resilience during the recovering process (CONSOL, 2013; European Commission, 2015). The high incidence of serious and fatal accidents involving elderly pedestrians highlights the need to not only reduce the number of elderly pedestrian accidents, but also to minimize the injury severity of these accidents.

This paper is divided into four sections. The first part provides the introduction; the second part reviews the state of the art on injury severity of traffic accidents and built environment; the third section describes the methodology, including data modelling and results; and the fourth section summarizes the findings and potential further research.

2. Injury severity of traffic accidents and built environment

Most studies on injury severity of vehicle-pedestrian collisions include factors of the accident, like time of the day, day of the week, season, and location; the pedestrian, like age and gender; the driver, like age and gender; and the vehicle, like type of vehicle. Researchers have also included on their analyses accident conditions (e.g., lightning, weather, and surface condition), depending on the quality of the traffic accident database. In addition, some authors have assessed the impact of built environment variables on injury severity, including socio-demographic features, such as population size (Pour-Rouholamin & Zhou, 2016) and population density (Clifton et al., 2009); land use features, such as land uses presence, land use mix and points of interest (POIs) (Casado-Sanz et al., 2019, 2020; Clifton et al., 2009; Kim et al., 2008, 2010; Park & Bae, 2020; Song et al., 2021); infrastructure features, such as intersections, crossings, traffic lights and bus stops per km (Clifton et al., 2009; Munira et al., 2020; Park & Bae, 2020; Rampinelli et al., 2022). This last approach provides valuable insight into how these features affect injury severity, but it requires collecting and processing data and linking it to each of the studied accidents. This process has been commonly conducted in the literature using two methods: collecting information from an area around each accident (Clifton et al., 2009; Park & Bae, 2020) or collecting information of territorial units (administrative or not) and assigning to each accident the information of the territorial unit where it occurred (Pour-Rouholamin & Zhou, 2016). This study used the second option due to the substantial number of accidents and geographical extension (Spain) under investigation, so each accident was linked to the built environment variables of the municipality where it was registered.

Regarding the methodology, researchers have used traditional classification models such as binary logistic regression (Casado-Sanz et al., 2019; Park & Bae, 2020), multinomial logistic regression (Casado-Sanz et al., 2020; Manner & Wünsch-Ziegler, 2013), ordered logistic regression (Mokhtarimousavi et al., 2020; Pour-Rouholamin & Zhou, 2016), and ordered probit regression (Clifton et al., 2009). These classification models are relatively easy to implement and interpret, but they only account for linear relationships between the independent variables and outcome. More recently, other classification machine learning (ML) techniques such as random forests (RF) or artificial neural networks (ANN) have been used to consider non-linearity (Li et al., 2017; Mokhtarimousavi et al., 2020). Although these algorithms are expected to outperform traditional models in classification scores, they might be more challenging to interpret, but the relative importance of the covariates can be analyzed (Friedman, 2001).

3. Methodology

3.1. Database

The objective of this study is to investigate the influence of accident and built environment factors on the injury severity of elderly pedestrian urban traffic accidents in Spain. To achieve this goal, two datasets were utilized: **(1)** accident features, including accident, pedestrian, driver, and vehicle information, and **(2)** Spanish municipalities built environment features, including socio-demographic, land use, and infrastructure information. These datasets were linked based on the municipality where each accident was registered.

The first dataset (1) was obtained from the National Database of Traffic Accidents (DGT) from 2016 to 2019. Out of 51,068 vehicle-pedestrian collisions during this period, 45,098 (88%) involved a single vehicle and single pedestrian configuration, which was the only configuration considered due to its high frequency. The accidents were grouped (Table 1) by pedestrian age (elderly or non-elderly) and injury level (fatality, serious injury, slight injury, or unharmed). Of the killed and seriously injured pedestrians, 71% and 41%, respectively, were elderly pedestrians, indicating their vulnerability, as they represent 28% of the pedestrian victims. The target variable was obtained by grouping the injury level of the pedestrian into two groups: slight accidents (slightly injured and unharmed), and severe accidents (killed and seriously injured). Accident, pedestrian, driver, and vehicle features were also collected for each accident (Table 2). Apart from these features, built environment influence should be assessed.

Table 1. Injury level of pedestrians involved in traffic crashes according to their age in Spain between 2016 and 2019.

Age of the pedestrian	Injury level of the pedestrian				Total
	Death	Seriously injured	Slightly injured	Unharmed	
Elderly	598 (71.1%)	2,238 (41.3%)	9,899 (25.8%)	54 (12.0%)	**12,789 (28.4%)**
Non-elderly	235 (27.9%)	3,076 (56.8%)	27,104 (70.6%)	332 (73.9%)	**30,747 (68.2%)**
Unknown	8 (1.0%)	99 (1.8%)	1,392 (3.6%)	63 (14.0%)	**1,562 (3.5%)**
Total	**841**	**5,413**	**38,395**	**449**	**45,098**

The second dataset (2) was obtained by processing data from multiple sources, including socio-demographic (INE), land use (SIOSE and OpenStreetMap (OSM)), and infrastructure (OSM) features. Socio-demographic data was assumed to be already representative of urbanized areas only. For the land use and infrastructure features, a systematic process was designed using GIS algorithms in ArcGIS (Esri Inc., 2020) and data wrangling in R (R Core Team, 2013) to obtain the features of the urbanized area of each municipality (provided by IGN), being Spain divided in more than 8,000 municipalities. The final list of built environment features is shown in Table 2.

The two datasets were joined based on the municipality where the accidents were registered. Samples with null or unspecified information were removed, resulting in a dataset of 35,646 accidents, 10,848 of which were suffered by elderly pedestrians. To test whether the effects of the variables on injury severity are specific for the elderly pedestrians or common for all pedestrians two samples were used: EP (elderly pedestrian accidents) and NEP (non-elderly pedestrian accidents) samples. The modelling of both samples is shown, interpreted, and compared in the following sections. Table 2 shows the collected variables considering the age and injury level of the pedestrians.

3.2. Logistic regression

In this study, a binary logistic regression (LR) was used to predict the injury level of the pedestrian who suffered the collision. As the target variable is a binary variable, the LR predicts the probability of an accident to be severe (seriously injured or killed pedestrian). The parameters were estimated using the maximum likelihood method.

3.3. Random Forest classifier

Random forest (RF) classifier is a supervised and ensemble learning method based on the decision tree classifier and bootstrap aggregating. The decision tree classifier is a classifier method based on decision tree learning, a method that, from the root node, splits the data considering the value of a particular feature until a “leaf”, a node that presents the output, is reached. Ensemble learning is based on constructing a strong learner from a set of weak predictors, and the prediction is obtained through the majority vote of the decision trees in this case. Bootstrap aggregating, or bagging, is an algorithm to avoid overfitting that consists of creating '*k*' subsamples of size '*m*' from a dataset of '*n*' samples with replacement and to train each of the '*k*' decision trees of the RF with each subsample.

Table 2. Summary statistics of the independent variables.

Variable	Category	EP sample: Elderly pedestrian accidents			NEP sample: Non-elderly pedestrian accidents		
		Total (N=10,848)	Slight (N=8,436) (77.8%)	Severe (N=2,412) (22.2%)	Total (N=24,798)	Slight (N=22,209) (89.6%)	Severe (N=2,589) (10.4%)
Accident characteristics							
Time of the day	Morning*	**4,096 (37.8%)**	3,156 (77.1%)	940 (22.9%)	**6,697 (27.0%)**	6,017 (89.8%)	680 (10.2%)
	Early morning	**41 (0.4%)**	32 (78.0%)	9 (22.0%)	**731 (2.9%)**	593 (81.1%)	138 (18.9%)
	Afternoon	**3,939 (36.3%)**	3,173 (80.6%)	766 (19.4%)	**8,787 (35.4%)**	7,974 (90.7%)	813 (9.3%)
	Evening	**2,253 (20.8%)**	1,696 (75.3%)	557 (24.7%)	**6,011 (24.2%)**	5,388 (89.6%)	623 (10.4%)
	Night	**519 (4.8%)**	379 (73.0%)	140 (27.0%)	**2,572 (10.4%)**	2,237 (87.0%)	335 (13.0%)
Weekday	Weekday*	**8,976 (82.7%)**	6,978 (77.7%)	1,998 (22.3%)	**20,270 (81.7%)**	18,221 (89.9%)	2,049 (10.1%)
	Weekend	**1,872 (17.3%)**	1,458 (77.9%)	414 (22.1%)	**4,528 (18.3%)**	3,988 (88.1%)	540 (11.9%)
Season	Spring*	**2,547 (23.5%)**	1,998 (78.4%)	549 (21.6%)	**6,455 (26.0%)**	5,744 (89.0%)	711 (11.0%)
	Summer	**2,538 (23.4%)**	2,010 (79.2%)	528 (20.8%)	**5,687 (22.9%)**	5,150 (90.6%)	537 (9.4%)
	Autumn	**2,863 (26.4%)**	2,227 (77.8%)	636 (22.2%)	**6,370 (25.7%)**	5,699 (89.5%)	671 (10.5%)
	Winter	**2,900 (26.7%)**	2,201 (75.9%)	699 (24.1%)	**6,286 (25.3%)**	5,616 (89.3%)	670 (10.7%)
Lightning	Daylight*	**8,833 (81.4%)**	6,991 (79.1%)	1,842 (20.9%)	**17,571 (70.9%)**	15,917 (90.6%)	1,654 (9.4%)
	Dawn dusk	**314 (2.9%)**	219 (69.7%)	95 (30.3%)	**859 (3.5%)**	765 (89.1%)	94 (10.9%)
	Artificial light	**1,545 (14.2%)**	1,104 (71.5%)	441 (28.5%)	**5,631 (22.7%)**	4,870 (86.5%)	761 (13.5%)
	Dark no light	**156 (1.4%)**	122 (78.2%)	34 (21.8%)	**737 (3.0%)**	657 (89.1%)	80 (10.9%)
Weather	Fair weather*	**10,049 (92.6%)**	7,861 (78.2%)	2,188 (21.8%)	**22,628 (91.2%)**	20,295 (89.7%)	2,333 (10.3%)
	Inclement weather	**799 (7.4%)**	575 (72.0%)	224 (28.0%)	**2,170 (8.8%)**	1,914 (88.2%)	256 (11.8%)
Surface	Dry*	**9,783 (90.2%)**	7,653 (78.2%)	2,130 (21.8%)	**22,124 (89.2%)**	19,840 (89.7%)	2,284 (10.3%)
	Surface condition	**1,065 (9.8%)**	783 (73.5%)	282 (26.5%)	**2,674 (10.8%)**	2,369 (88.6%)	305 (11.4%)
Location of the accident	Road section*	**6,295 (58.0%)**	4,837 (76.8%)	1,458 (23.2%)	**15,876 (64.0%)**	14,148 (89.1%)	1,728 (10.9%)
	Road intersection	**4,553 (42.0%)**	3,599 (79.0%)	954 (21.0%)	**8,922 (36.0%)**	8,061 (90.3%)	861 (9.7%)
Pedestrian characteristics							
Age	(numeric)**	**76.5 (7.36)**	76.1 (7.32)	77.7 (7.38)	**34.7 (18.2)**	34.4 (18.0)	37.7 (19.1)
Gender	Man*	**4,817 (44.4%)**	3,718 (77.2%)	1,099 (22.8%)	**11,391 (45.9%)**	10,108 (88.7%)	1,283 (11.3%)
	Woman	**6,031 (55.6%)**	4,718 (78.2%)	1,313 (21.8%)	**13,407 (54.1%)**	12,101 (90.3%)	1,306 (9.7%)
Driver characteristics							
Age	(numeric)**	**51.9 (77.6)**	52.9 (83.1)	48.7 (53.9)	**53.5 (93.7)**	54.2 (96.9)	47.5 (58.5)
Gender	Man*	**8,208 (75.7%)**	6,353 (77.4%)	1,855 (22.6%)	**18,275 (73.7%)**	16,289 (89.1%)	1,986 (10.9%)
	Woman	**2,640 (24.3%)**	2,083 (78.9%)	557 (21.1%)	**6,523 (26.3%)**	5,920 (90.8%)	603 (9.2%)
Vehicle characteristics							
Vehicle type	Car*	**8,858 (81.7%)**	6,906 (78.0%)	1,952 (22.0%)	**20,231 (81.6%)**	18,155 (89.7%)	2,076 (10.3%)
	Motorcycle	**803 (7.4%)**	606 (75.5%)	197 (24.5%)	**2,436 (9.8%)**	2,188 (89.8%)	248 (10.2%)
	Bus	**184 (1.7%)**	118 (64.1%)	66 (35.9%)	**446 (1.8%)**	332 (74.4%)	114 (25.6%)
	Truck	**372 (3.4%)**	262 (70.4%)	110 (29.6%)	**485 (2.0%)**	396 (81.6%)	89 (18.4%)
	Bicycle	**502 (4.6%)**	451 (89.8%)	51 (10.2%)	**917 (3.7%)**	883 (96.3%)	34 (3.7%)
	Other	**129 (1.2%)**	93 (72.1%)	36 (27.9%)	**283 (1.1%)**	255 (90.1%)	28 (9.9%)
Municipality characteristics							
Socio-demography							
Population size (inhabitants)	< 10,000*	**857 (7.9%)**	555 (64.8%)	302 (35.2%)	**1,096 (4.4%)**	913 (83.3%)	183 (16.7%)
	10,000 to 50,000	**2,409 (22.2%)**	1,804 (74.9%)	605 (25.1%)	**4,737 (19.1%)**	4,233 (89.4%)	504 (10.6%)
	50,000 to 100,000	**1,856 (17.1%)**	1,421 (76.6%)	435 (23.4%)	**4,802 (19.4%)**	4,296 (89.5%)	506 (10.5%)
	100,000 to 500,000	**3,675 (33.9%)**	2,933 (79.8%)	742 (20.2%)	**8,661 (34.9%)**	7,749 (89.5%)	912 (10.5%)
	> 500,000	**2,051 (18.9%)**	1,723 (84.0%)	328 (16.0%)	**5,502 (22.2%)**	5,018 (91.2%)	484 (8.8%)
Population density	(numeric)**	**107.0 (74.5)**	111.0 (76.3)	92.2 (65.9)	**115.0 (75.7)**	116.0 (76.6)	101.0 (65.8)
Main house rate	(numeric)**	**0.775 (0.12)**	0.780 (0.12)	0.755 (0.12)	**0.795 (0.11)**	0.797 (0.11)	0.776 (0.12)
Land use							
Primary proportion	(numeric)**	**0.110 (0.14)**	0.104 (0.13)	0.131 (0.15)	**0.092 (0.12)**	0.090 (0.12)	0.106 (0.12)
Secondary proportion	(numeric)**	**0.055 (0.06)**	0.056 (0.06)	0.052 (0.06)	**0.057 (0.07)**	0.058 (0.07)	0.054 (0.06)
Tertiary proportion	(numeric)**	**0.137 (0.08)**	0.141 (0.08)	0.126 (0.08)	**0.147 (0.07)**	0.148 (0.07)	0.144 (0.08)
Residential proportion	(numeric)**	**0.518 (0.14)**	0.520 (0.14)	0.509 (0.14)	**0.523 (0.14)**	0.524 (0.14)	0.513 (0.13)
Land use mix	(numeric)**	**0.646 (0.10)**	0.645 (0.10)	0.651 (0.10)	**0.646 (0.10)**	0.645 (0.10)	0.656 (0.10)
POIs per km	(numeric)**	**6.10 (6.09)**	6.42 (6.29)	5.00 (5.20)	**6.32 (6.23)**	6.40 (6.30)	5.63 (5.59)
Infrastructure							
Road intersections per km	(numeric)**	**8.78 (1.33)**	8.77 (1.31)	8.81 (1.40)	**8.76 (1.32)**	8.76 (1.32)	8.78 (1.32)
Crossings per km	(numeric)**	**4.97 (4.25)**	5.19 (4.37)	4.19 (3.71)	**5.34 (4.31)**	5.42 (4.37)	4.62 (3.74)
Traffic signals per km	(numeric)**	**0.89 (1.25)**	0.97 (1.31)	0.61 (0.96)	**0.97 (1.33)**	1.00 (1.35)	0.73 (1.06)
Bus stops per km	(numeric)**	**0.65 (0.52)**	0.68 (0.53)	0.55 (0.48)	**0.69 (0.50)**	0.69 (0.50)	0.64 (0.46)

* Baseline category for the logistic regression model

** For numeric variables the mean and standard deviation are displayed [mean (SD)]

RF classifier is expected to outperform the accuracy of more traditional methods, such as logistic regression, as it can consider non-linear relations between the covariates and the target feature. In contrast, the functioning of the RF is hard to understand due to the large number of trees and their complexity. Understanding the contribution of the covariates is difficult, but a ranking of the variable importance can be obtained using two values: the mean decrease accuracy (MDA) and the mean decrease impurity (MDI). The MDA of a variable is the rate of change of the accuracy when that variable is randomly permuted, using the out-of-bag subsample in each tree. If permuted, more important variables will affect MDA more than less important ones, so a high MDA represents a high usefulness of a variable in the classification. The MDI is the average contribution of each variable to the homogeneity of the nodes, measured by the Gini impurity, in the splitting process over all the trees of the RF. Features that contribute more to raising the homogeneity of the nodes (higher MDI) are more important in the classification.

3.4. Models training

3.4.1. Train and test sets

To avoid over-optimistic results of the trained ML algorithms, they should be trained and tested with different data. The train set is commonly used to train the models and perform hyperparameters optimization, while the test set is only used for testing the final models. In this study, the train and test sets used were obtained from the elderly and non-elderly accidents datasets by randomly extracting 80% and 20% of the samples, respectively. Furthermore, this process was set to maintain the distribution of the target variable from the original databases.

3.4.2. Imbalanced data classification

Working with imbalanced data is a common issue in classification problems using real-world data, and this is the case of our study, as slight accidents represent about 80% and 90% (ratios of 1:4 and 1:9) of the elderly and non-elderly pedestrian accidents samples, respectively. The nature of the original datasets might affect the training process of the ML algorithms, and these algorithms may overfit and generalize the majority class. Road safety injury severity studies normally do not account for this issue (Fiorentini & Losa, 2020), and researchers usually give an overall result for the model (accuracy), which tends to be high as the algorithm classifies properly the majority class.

To deal with imbalanced data, resampling techniques (undersampling and oversampling) should be used to train the model with balanced data. On the one hand, **undersampling techniques** reduce the number of samples of the majority class by sampling '*n*' samples of the majority class, being '*n*' the number of samples of the minority class. This method reduces the computing cost and uses real data only, but potentially useful data is removed. On the other hand, **oversampling techniques** augment the number of minority class samples until there is a 50/50 balance, by duplicating samples or creating new samples of the minority class. Hence, the ML algorithm is trained with synthetic data and this algorithm might lead to overfitting if the new samples are duplicates of the real data. As both techniques have advantages and drawbacks, both of them were tested and compared with models trained without resampling. The chosen **undersampling technique** is random undersampling and the **oversampling technique** is 'SMOTE' (Chawla et al., 2002), implemented through 'SMOTE-NC' class of Python's 'imbalanced-learn' package (Lemaître et al., 2017), as it is designed to work with datasets with both categorical and numerical features.

Table 3. F_1 score, AUC, and accuracy of the logistic regression and random forest algorithms for the elderly and non-elderly pedestrian accidents.

Model	Resampling technique	EP sample: Elderly pedestrian accidents				NEP sample: Non-elderly pedestrian accidents			
		Hyper-parameters*	F_1 score	AUC	Accuracy	Hyper-parameters*	F_1 score	AUC	Accuracy
Logistic regression (LR)	*Base dataset*	-	0.024	0.619	0.778	-	0.004	0.620	0.896
	Undersampling	-	**0.382**	0.617	0.589	-	**0.229**	0.614	0.605
	Oversampling	-	0.340	0.565	0.583	-	0.180	0.527	0.620
Random forest (RF)	*Base dataset*	100, 30, 1	0.161	0.635	0.770	100, 30, 1	0.029	0.671	0.894
	Undersampling	300, 10, 2	**0.425**	0.662	0.608	300, 50, 10	**0.247**	0.664	0.605
	Oversampling	300, 5, 10	0.371	0.613	0.608	500, 5, 1	0.189	0.567	0.631

* Hyperparameters for the random forest are: [number of tress], [maximum depth], [minimum samples in a leaf node]

3.4.3. Hyperparameter optimization: k-fold cross-validation

ML algorithms usually have several inputs that can be divided in two groups: training data and hyperparameters). Variation of the hyperparameters might affect the accuracy, so obtaining the set of hyperparameters that presents the best classification is needed. This process is known as **hyperparameter optimization**, or hyperparameter tuning.

In this case, *k*-fold cross-validation technique was used, which consists of dividing the training set in '*k*' folds, being 5 or 10 usual values for '*k*', and training '*k*' models by using *k*-1 subsets for the training and 1 for the testing. In this research, an exhaustive hyperparameter optimization was carried out using a 10-fold cross-validation.

Regarding the hyperparameter to optimize; on the one hand, **logistic regression** has some hyperparameters to tune, but performance might not be affected as these are more related with computing cost, so hyperparameters were set by default. On the other hand, **random forest** classifier has several hyperparameters related to the algorithm's topology, and 3 hyperparameters were tuned testing the following values: the number of decision trees of the forest [100, 300, 500, 800, 1000, 1500], the maximum number of levels in each decision tree [5, 10, 20, 30, 50], and the minimum number of data samples allowed in a leaf node [1, 2, 5, 10]. Note that, as different resampling techniques will be applied to balance the output feature of the dataset to train the model, these should also be applied during the cross-validation process and not before it, so the fold used to test in each iteration is not affected by the resampling.

3.5. Assessment of the models

Both trained LR and RF models were compared based on metric scores. Normally, the performance of a classifier is assessed using the accuracy, but this might lead to over-optimistic results if the dataset is imbalanced as this score would provide advantage to the majority class (López et al., 2013). In consequence, other metrics that consider class imbalance should be used, and in this paper the F_1 score (harmonic mean of recall and precision) and area under the curve (AUC) were used. Table 3 shows the F_1 score, AUC, and accuracy of the best models after the hyperparameter optimization with the base, undersampled and oversampled datasets. Considering the F_1 score, the **models with the best performance** were the LR and RF trained with the undersampled datasets for both EP and NEP samples, presenting the RFs the best performance. These models were chosen for further interpretation in the next section.

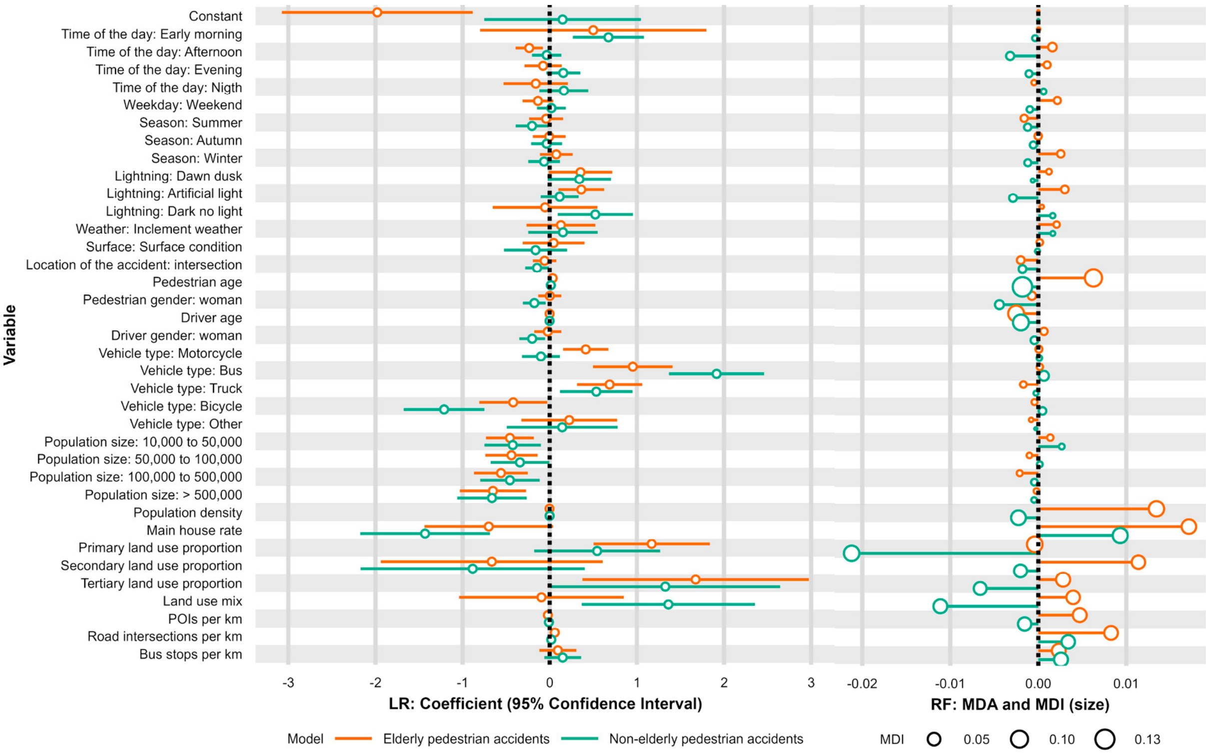
3.6. Modelling results

This section presents the interpretation of the logistic regression (LR) and random forest (RF) models with the best metric scores, as described in the previous section, which was carried out by analyzing the parameters and *p*-values of the LR, and the relative importance of the independent variables (MDA and MDI) for the RF. Figure 1 displays graphical representations of the parameters of the LR models and their 95% confidence intervals, as well as the MDI and MDA of the variables in the RF models.

Regarding the LR, elderly and non-elderly pedestrian accidents were affected by features of all groups. **For all pedestrian accidents**, it was observed that pedestrians were more likely to suffer severe accidents if they were older, the accident occurred at dawn or dusk, the municipality had high tertiary land use proportion, or if they are hit by a bus or truck, compared to a car collision. Conversely, accidents were less severe if the driver was older, or if the vehicle was a bicycle. Interestingly, the effect of vehicle type on the accident injury severity was higher for non-elderly pedestrians, which could be due to the intrinsic fragility of older people, who might suffer more severe accidents regardless of the vehicle involved in the collision. In addition, populations above 10,000 inhabitants expected less severe pedestrian accidents, and this effect was greater for larger cities, being those with more than 500,000 inhabitants the safest. **In particular, elderly pedestrians** were also more likely to suffer slight accidents during the afternoon, compared to morning hours, in denser populated municipalities with high values of POIs per km, while severe accidents were expected in dark condition with artificial lightning, motorcycle collisions, and in municipalities with high primary land use and intersections per km values. Regarding lightning, elderly pedestrians being negatively affected by artificially lighted locations, while this effect is not significant for the rest of pedestrians (although the coefficient indicates a contrary effect), suggests that these spots might not be properly lighted for elder people. **On the other hand, non-elderly pedestrians** were expected to suffer more severe accidents in the early morning, in dark and no-lighted areas, and in municipalities with high land use mix, and less severe accidents during summer, at road intersections, compared to accidents occurred at road sections, and if the driver and pedestrian were women. Accidents at intersections were less severe, which could be explained by vehicle speed reduction in these locations, being speed a key factor to explain the injury level of the pedestrian (not considered in this study because of lack of data); the injury level of the elderly pedestrian accidents not being affected by this fact indicates that this speed reduction is not sufficient to reduce the accident's severity compared to the intrinsic fragility of this age group.

About the RF, it was surprising to find that the municipality variables had the greatest impact on the RF models' results, with the main house rate and road intersections per km being the most important variables for both samples. Important features for the RF that were also statistically significant in the LR for both EP and NEP samples are the main house rate and municipalities with a population size of 10,000 to 50,000 inhabitants, while the rest of studied population size groups do not show high MDA values, being even negative in some cases. Additionally, the LR and RF of the elderly pedestrian sample are affected if the accidents occurred during the afternoon, at dawn or dusk, or in dark spots artificially lighted, being also affected by the population density, tertiary land use proportion, POIs per km and road intersections per km. Other features that showed relatively high MDA values were not significant in the LR included evening, weekend, winter, secondary land use proportion and land use mix for the elderly pedestrian sample; and inclement weather and bus stops per km for both samples. Interestingly, although the age of the pedestrian was significant for all pedestrians in the LR, it was only relatively important for the elderly pedestrian accidents, which highlights the need to study elderly pedestrian accidents in more detail by sub-groups of pedestrian's age (e.g., 65-74, 75-84, 85-94 and 95+).

Figure 1. Coefficient (95% Confidence Interval) of the logistic regressions and MDA and MDI of the random forests.



4. Conclusions

This study utilized logistic regression and random forest models to analyze relationship between the injury severity of vehicle collisions suffered by elderly pedestrians in Spain from 2016 to 2019, and accident and built environment features. The injury severity of non-elderly pedestrian accidents was also assessed to determine if the effects of the variables were specific to elderly pedestrians or applicable to all pedestrians. For this purpose, an ad hoc dataset was built with accident and built environment variables of the municipality where each accident was registered, and accidents were grouped into severe and slight accidents. Resampling techniques were used to train the logistic regression (LR) and random forest (RF) models, with the models trained with the undersampled datasets achieving the best metric scores. RF outperformed LR performance, but the LR results were easier to interpret.

The findings highlight that, although in smaller towns less accidents are registered in absolute terms, these areas should be given special attention as pedestrian accidents showed to be more severe than in larger cities for all pedestrians. Furthermore, these smaller cities usually present low main house rates, which is linked to more severe pedestrian accidents; and low values of population density and POIs per km, and high figures of primary land use and road intersections per km, which are linked to more severe elderly pedestrian accidents. With respect to lightning conditions, elderly pedestrian accidents are more severe in dark with artificial lights condition; which suggests that, although elderly pedestrians are more cautious, these spots are not properly lighted during dark hours for them, potentially due to elderly individuals wearing dark clothes or having visually impairments; hence, lightning devices should be improved with these considerations. Interestingly, the age of the pedestrian affects both LR and RF only for the elderly pedestrian sample, which indicates the need to study this age group in more detail.

This study provides valuable insight in the elderly pedestrian road safety, but further research is needed. RF algorithm outperformed LR, but additional investigation is required to interpret the results. Additionally, other ML and variable selection techniques should be tested tuning all possible hyperparameters using resampling techniques.

Acknowledgements

Daniel Gálvez-Pérez is developing his doctoral thesis while he enjoys a grant from to the Universidad Politécnica de Madrid through the 'Programa Propio de I + D + I 2020: Ayudas para Contratos Predoctorales'.

References

Casado-Sanz, N., Guirao, B., & Attard, M. (2020). Analysis of the risk factors affecting the severity of traffic accidents on Spanish crosstown roads: The driver's perspective. *Sustainability*, *12*(6), 2237.
Casado-Sanz, N., Guirao, B., & Gálvez-Pérez, D. (2019). Population ageing and rural road accidents: Analysis of accident severity in traffic crashes with older pedestrians on Spanish crosstown roads. *Research in Transportation Business & Management*, *30*, 100377.
Chawla, N. V, Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357.
Clifton, K. J., Burnier, C. v, & Akar, G. (2009). Severity of injury resulting from pedestrian–vehicle crashes: What can we learn from examining the built environment? *Transportation Research Part D: Transport and Environment*, *14*(6), 425–436.
CONSOL. (2013). *Mobility Patterns in the Ageing Populations (Final technical report of WP2 of the 7th framework EC project CONSOL)*.
DGT. (2020). *Tendencias de la movilidad y siniestralidad en vías urbanas*.
Esri Inc. (2020). *ArcMap* (10.8.1).
European Commission. (2015). *ElderSafe-Risks and countermeasures for road traffic of elderly in Europe (No. MOVE/C4/2014-244)*. European Commission–Directorate-General for Mobility and Transport (DG-MOVE).
European Commission. (2021a). *Road safety thematic report - Seniors. European Road Safety Observatory*.
European Commission. (2021b). *Facts and Figures Seniors. European Road Safety Observatory*.
Fiorentini, N., & Losa, M. (2020). Handling imbalanced data in road crash severity prediction by machine learning algorithms. *Infrastructures*, *5*(7), 61.
Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 1189–1232.
Kim, J.-K., Ulfarsson, G. F., Shankar, V. N., & Kim, S. (2008). Age and pedestrian injury severity in motor-vehicle crashes: A heteroskedastic logit analysis. *Accident Analysis & Prevention*, *40*(5), 1695–1702.
Kim, J.-K., Ulfarsson, G. F., Shankar, V. N., & Mannering, F. L. (2010). A note on modeling pedestrian-injury severity in motor-vehicle crashes with the mixed logit model. *Accident Analysis & Prevention*, *42*(6), 1751–1758.
Lemaître, G., Nogueira, F., & Aridas, C. K. (2017). Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *The Journal of Machine Learning Research*, *18*(1), 559–563.
Li, D., Ranjitkar, P., Zhao, Y., Yi, H., & Rashidi, S. (2017). Analyzing pedestrian crash injury severity under different weather conditions. *Traffic Injury Prevention*, *18*(4), 427–430.
López, V., Fernández, A., García, S., Palade, V., & Herrera, F. (2013). An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information Sciences*, *250*, 113–141.
Manner, H., & Wünsch-Ziegler, L. (2013). Analyzing the severity of accidents on the German Autobahn. *Accident Analysis & Prevention*, *57*, 40–48.
Mokhtarimousavi, S., Anderson, J. C., Azizinamini, A., & Hadi, M. (2020). Factors affecting injury severity in vehicle-pedestrian crashes: A day-of-week analysis using random parameter ordered response models and Artificial Neural Networks. *International Journal of Transportation Science and Technology*, *9*(2), 100–115.
Munira, S., Sener, I. N., & Dai, B. (2020). A Bayesian spatial Poisson-lognormal model to examine pedestrian crash severity at signalized intersections. *Accident Analysis & Prevention*, *144*, 105679.
Park, S.-H., & Bae, M.-K. (2020). Exploring the determinants of the severity of pedestrian injuries by pedestrian age: a case study of Daegu Metropolitan City, South Korea. *International Journal of Environmental Research and Public Health*, *17*(7), 2358.
Pour-Rouholamin, M., & Zhou, H. (2016). Investigating the risk factors associated with pedestrian injury severity in Illinois. *Journal of Safety Research*, *57*, 9–17.
R Core Team. (2013). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna*.
Rampinelli, A., Calderón, J. F., Blazquez, C. A., Sauer-Brand, K., Hamann, N., & Nazif-Munoz, J. I. (2022). Investigating the risk factors associated with injury severity in pedestrian crashes in Santiago, Chile. *International Journal of Environmental Research and Public Health*, *19*(17), 11126.
Song, L., Fan, W. D., Li, Y., & Wu, P. (2021). Exploring pedestrian injury severities at pedestrian-vehicle crash hotspots with an annual upward trend: A spatiotemporal analysis with latent class random parameter approach. *Journal of Safety Research*, *76*, 184–196.