# Evaluation of Pedestrian Crossing Accidents Using Artificial Neural Network

Bertha Santos[1(✉)], Jorge Gonçalves[2], Shohel Amin[3], Sandra Vieira[4], and Carlos Lopes[5]

[1] University of Beira Interior, GeoBioTec and CERIS, 6200-358 Covilhã, Portugal
bsantos@ubi.pt

[2] University of Beira Interior, GeoBioTec and CITTA, 6200-358 Covilhã, Portugal
jorge@ubi.pt

[3] Coventry University, Coventry CV1 5FB, UK
s.amin15@aston.ac.uk

[4] National Laboratory for Civil Engineering, 1700-066 Lisboa, Portugal
sandravieira@lnec.pt

[5] National Road Safety Authority, 2734-507 Barcarena, Portugal
carlos.lopes@ansr.pt

**Abstract.** Most of European cities face increasing problems caused by excessive traffic of conventional fuel-based transport modes. To reverse this situation, sustainable urban mobility policies have been promoting soft modes of transport, such as walking. Despite the advantages of walking in reducing traffic congestion and pollution, cities have not always evolved to accommodate the needs of pedestrians. According to the European Commission, in 2020, 20% of road fatalities in the European Union (EU) and 21% in Portugal were pedestrian. Pedestrian fatality rates per million population was 9.7 for all EU countries and 13.1 for Portugal. In European and Portuguese urban areas, 36% and 27% of the fatalities were pedestrians' and 49% and 56% of all pedestrian fatalities were elderly's (respectively). In pedestrian infrastructures, crossings are considered the most critical element due to conflicts between vehicles and pedestrians. It is then essential to identify and minimize risk factors that increase the probability of accidents in these locations. The proposed work intends to assess this challenge by using Artificial Neural Network (ANN) to create pedestrian severity prediction models and identify road and pedestrian risk factors for accident occurred in or near urban crossings. The official Portuguese database on run over pedestrian accidents occurred between 2017–2021 was analyzed with ANN considering two scenarios: pre-Covid-19 and during Covid-19 period. Results obtained demonstrate that the use of ANN can promote a proactive infrastructure management, suggesting that crossings traffic lights operation, lighting, shoulders and pavement conditions, high speed limits (51–90 km/h) and pedestrians moving in soft modes are critical factors.

**Keywords:** Road Safety · Pedestrian Accidents at Urban Crossings · Artificial Neural Network (ANN) · Severity Predictive Model · Risk Factors


C. McNally et al. (Eds.): TRAconference 2024, LNMOB, pp. 717–725, 2026.
https://doi.org/10.1007/978-3-031-88974-5_103

# 1 Introduction

## 1.1 Framework and Objectives

Despite the advantages of walking to reduce traffic congestion and pollution, cities have not always evolved to accommodate the needs of pedestrians and walking has, in many cases, been neglected in the development of transport systems [1]. Due to their vulnerability, accidents involving pedestrians show higher severity when compared with other kinds of accidents, with a significant rate of morbidity, disability, and death. Estimates suggest that approximately 12 million road accidents involving pedestrians occur every year worldwide and cause the deaths of about 270000 people (around 23% of all traffic fatalities) [2]. Between 2010 and 2020, the absolute number of pedestrian fatalities fell 39% in the EU27 and 69% in Portugal. The number of road fatalities in the same period decreased 32% in the EU27 and 43% in Portugal [3–5]. Therefore, the proportion of pedestrians in the number of road fatalities decreased in the last decade 11% in the EU27 and 9% in Portugal. In 2020, 20% of all traffic fatalities were pedestrian and pedestrian fatality rate per million population was 9.7 in the EU. For Portugal, these numbers were 21% and 13.1, respectively [6, 7]. For the same year, inside EU and Portuguese urban areas, 36% and 27% of the fatalities were pedestrians' and 49% and 56% of all pedestrian fatalities were elderly's (>65 years) (respectively) [5, 6]. Therefore, an understanding of the risk factors affecting pedestrians' safety is essential to define effective measures to protect pedestrians. Quantitative safety assessment is generally preformed through Accident Prediction Models (APM). Several Pedestrian APM have been applied to roads sections [2, 8], intersections [9–15] and crossings [14, 15]. In view of the above, this study aims to produce pedestrian severity prediction models and identify road and pedestrian risk factors for accident occurred in or near urban crossings using Artificial Neural Network. The pre-Covid19 and Covid19 periods are considered separately to assess whether the restrictions imposed on the movement of vehicles and people, during the pandemic, modified the risk factors, number and severity of pedestrian crossing accidents. The methodology was applied to the 2017–2021 ANSR (Portuguese National Road Safety Authority) database.

## 1.2 Pedestrian Accidents

Recently two scientific papers were published with an extensive bibliometric review about pedestrian safety studies [16, 17]. In one of the studies, Mirhashemi et al. (2022) [16] identified 7 clusters for highly cited papers about pedestrian safety: pedestrian accident frequency models, pedestrian injury severity models, traffic engineering measures in pedestrians' safety, global reports around pedestrian accident epidemiology, effect of age and gender on pedestrians' behavior, pedestrian crowd dynamics and evacuation and distraction of pedestrians. When choosing the best modeling approach, the following issues were considered: over-dispersion, under-dispersion, time-varying explanatory variables, temporal and spatial correlation, low sample-mean and small sample size, injury-severity and crash-type correlation, under-reporting, omitted-variables bias, endogenous variables, functional form, and fixed parameters. In the second study, Shrinivas et al. (2023) [17] performed a systematic analysis of 93 articles about the impact of variables on the severity of pedestrian injuries. They examined a total of 904655 pedestrian incidents from 19 different countries and categorized risk factors in 7 areas: pedestrian features (age, gender and race), driver's features (age and gender), vehicle features (vehicle type and speed), road and environmental features (weather, road type, traffic control devices and lighting condition), accident features (collision type and pre-crash movements), built environment (network connectivity, accessibility, and land use), and spatial and temporal patterns. In total, 204 factors were examined. When modelling, the authors suggest considering 15 variables: road environment (urban/rural), driver gender, driver intoxicated, hit and run, land use, lighting, maneuver, number of lanes, pedestrian age, pedestrian behavior, pedestrian gender, pedestrian intoxicated, road geometry, road type, and speed limits. Concerning methodologies, recent studies have used Multiple Correspondence Analysis (MCA), random parameter logit and probit models, and Machine Learning (ML) techniques such as Artificial Neural Network (ANN), Support Vector Machine (SVM), Boosting algorithms or Bayesian models [16].

### 1.3 Machine Learning Applied to Accident Data

Accident analysis is complex due to the presence of human behavior, which is difficult to predict and model due to the high subjective variations in people's decision making. Traditionally, statistical modeling techniques have been used to predict accidents and classify their severity: Poisson regression, binomial regression, negative binomial regression, Poisson-lognormal regression, gamma regression, zero-inflated regression, generalized estimation equations, negative multinomial model, random effects model and random parameters model. However, the limitations of these approaches have been widely explored, offering an opportunity to use new approaches, such as ML techniques [18]. ML can be divided into supervised, unsupervised, and semi-supervised learning. In supervised learning, the response values of the cases in the training set are known and the model response to the current pattern of inputs is evaluated, allowing changes that bring the model response closer to the expected (known) response. Data accident can be considered in this category.

Theoretically, given a set of input data ($x_i$, $y_i$), where $x$ is the vector of the variables that represent the predictive attributes and $y$ is the label of the class that the case belongs to, the model's task is to learn a target function $f$ that maps each set of attribute $x$ in one of the predefined $y$ classes or label (attribute that describes a characteristic of the phenomenon of interest, in this case, pedestrian injury severity). As in pedestrian injury severity analysis the class is discrete, it is a classification problem. Despite the use of ML techniques in a variety of road safety problems, examples of its application to pedestrian accidents are still limited. Some examples include bayesian model [9], association rules mining [13], gradient boosting [2] and ANN [10–12, 19].

## 2 Method

### 2.1 Accident Data

The pedestrian accident data analyzed in this study was provided by the Portuguese National Authority of Road Safety (ANSR). The information concerns 24774 pedestrians involved in 24211 accidents with victims (fatal, serious, and minor victims) that occurred in mainland Portugal and the Azores and Madeira islands, between 2017 and 2021 (5 years). The data was split in two datasets to evaluate the effect of the Covid19 pandemic on number, severity and risk factors associated with this type of accidents.

### 2.2 Variables and Descriptive Statistics

The data preparation process for descriptive analysis and subsequent modeling included the following main steps: selection of run over pedestrian accidents; data organized by pedestrian involved in accident (one line by pedestrian); selection of accidents in urban environment and urban streets coded as 'pedestrian hit' ('run over and run away' and 'animal hit' accidents were eliminated); selection of relevant pedestrian actions: 'crossing a marked crossing', 'crossing a marked crossing without respecting traffic lights' and 'crossing outside a pedestrian crossing, at less than 50 m from a crossing'; definition of two injury severity groups: unharmed+minor and serious+fatal; based on the local speed limits, definition of speed classes: 0–30, 31–50 and 51–90 km/h; database division in two datasets: Pre-Covid19 (from 1/January/2017 to 17/March/2020) and Covid19 (from 18/march/2020 (first confinement) to 31/December/2021); consideration of 3 hourly time periods: morning and afternoon peak periods (8–10 h and 17–19 h) and intermediate period (between peaks, i.e., 11–16 h). Other variable data classes were created based on the distribution of data along the days of the week and months of the year, according to weather condition, pavement grip condition, vertical design, circulation direction, type of intersection, vertical signs, local speed and pedestrian age.

The evaluation of the variables' percentages revealed that in general, there were no significant changes between the Pre-Covid19 and Covid19 periods (less than $\pm$ 2%). However, an increase of 1.3% in the number of pedestrians with serious and fatal injuries was verified. For the independent variables, the most significant variations were the number of pedestrians involved in accidents occurred between 11am and 4pm (+2.9%), during the daytime (+6.4%), with a similar decrease at nighttime with lighting (−6.9%), clear weather conditions (+4.6%) and rain (−4.5%). There was also an increase in pedestrians involved in accidents occurred in wet or humid pavement (+5.1%) and a decrease for pavements in good condition (-4.8%). The monthly variation was not analyzed due to the different datasets time span.

To produce the predictive models for pedestrian severity and evaluate the importance of the predictive variables, the Multilayer Perceptron (MLP) procedure available in the Neural Network option of IBM SPSS Statistics 28.0.0.0 is used. The aim is to produce 3 prediction models for each dataset (Pre-Covid19 and Covid19): a complete model with all the considered predictive variables, a model for 'road characteristics' variables and a model for 'pedestrian' variables. In the modelling process, 70% of the dataset is used for training and 30% for testing, and the automatic architecture option is selected to build a network with one hidden layer. To minimize the total error, the Batch type of training with Scaled conjugate gradient optimization algorithm (to estimate the synaptic weights) is considered. The default SPSS training options are adopted. Tables 1 and 2 present the main outputs of the Pre-Covid19 and Covid 19 models. Several analyses were conducted with all the independent variable classes and using only the variables selected as aggravate/contributing condition for pedestrian accident occurrence. The two approaches revealed similar results, so only the models generated

**Table 1.** Pre-Covid19 models

| Model | Complete | Road characteristics | Pedestrian |
|---|---|---|---|
| **Training**<br>% incorrect predictions<br>% correct classification | 7.2%<br>92.7%* | 7.3%<br>92.7%* | 7.1%<br>92.9%* |
| **Testing**<br>% incorrect predictions<br>% correct classification | 7.3%<br>92.7%* | 7.1%<br>92.9%* | 7.4%<br>92.6%* |
| **ROC**<br>Area under the curve | 0.668 for class 1<br>0.668 for class 2 | 0.567 for class 1<br>0.567 for class 2 | 0.647 for class 1<br>0.647 for class 2 |
| **Top 10 Independent variable importance** | LSp: 51–90 km/h<br>PT: Cement<br>LS: Off<br>T: Moving on skate, scooter or others<br>PC: Poor<br>S: Unpaved shoulder<br>PT: Stone<br>L: Night, no lighting<br>VS: Giving way<br>LSp: 31–50 km/h | LSp: 51–90 km/h<br>VS: Stop<br>PT: Stone<br>PT: Cement<br>S: Unpaved shoulder<br>LSp: 31–50 km/h<br>HS: With marks separating direction/lanes of traffic<br>LSp: 0–30 km/h<br>LS: Intermittent<br>PC: Poor | A: > 65 years old T: Pushing a bicycle, Stroller or people with Physical disabilities<br>A: 50–64 years old<br>T: Pedestrian in group<br>T: Moving on skate, scooter or others<br>PA: Crossing a marked crossing without respecting traffic lights<br>A: 0–9 years old<br>A: 10–17 years old<br>T: Isolated pedestrian<br>PA: Crossing outside a pedestrian crossing, at less than 50m from a crossing |

Number of cases = 9568 (92.8% Unharmed+Minor, 7.2% Serious+Fatal)
Dependent variable: Injury severity (class 1 = Unharmed+Minor, class 2 = Serious+Fatal)
Independent variables: Temporal (Hour H, Day of the week DW, Month M), Road environment (Luminosity L, Weather W), Road characteristics (Horizontal design HD, Vertical design VD, Shoulder S, Circulation direction CD, Intersection I, Vertical signs VS, Horizontal signs HS, Light signs LS, Pavement type PT, Pavement grip PG, Pavement condition PC, Local speed LSp), Pedestrian (Age A, Sex, Type T, Pedestrian action PA)
*All cases of the dependent variable classified in class 1

**Table 2.** Covid19 models

| Model | Complete | Road characteristics | Pedestrian |
|---|---|---|---|
| **Training**<br>% incorrect predictions<br>% correct classification | 8.6%<br>91.4%* | 7.8%<br>92.2%* | 8.5%<br>91.5%* |
| **Testing**<br>% incorrect predictions<br>% correct classification | 8.2%<br>91.8%* | 9.9%<br>90.1%* | 8.4%<br>91.6%* |
| **ROC**<br>Area under the curve | 0.630 for class 1<br>0.630 for class 2 | 0.604 for class 1<br>0.604 for class 2 | 0.626 for class 1<br>0.626 for class 2 |
| **Top 10 Independent variable importance** | T: Moving on skate, scooter or others<br>VS: Stop<br>LS: Off<br>LSp: 51–90 km/h<br>L: Blinding sun<br>S: Unpaved shoulder<br>PT: Cement<br>PC: Poor<br>L: Night, no lighting<br>LS: Intermittent | PC: Poor<br>LSp: 0–30 km/h<br>I: Roundabout<br>LSp: 51–90 km/h<br>PT: Stone<br>PT: Asphalt<br>LS: Intermittent<br>I: 4-leg intersection<br>S: Unpaved shoulder<br>VS: Stop | T: Pushing a bicycle, stroller or people with physical disabilities<br>A: 18–24 years old<br>PA: Crossing a marked crossing without respecting traffic lights<br>T: Pedestrian in group<br>T: Moving on skate, scooter or others<br>PA: Crossing a marked crossing<br>A: 10–17 years old<br>A: 25–49 years old<br>T: Isolated pedestrian<br>A: 0–9 years old |

Number of cases = 3310 (91.5% Unharmed+Minor, 8.5% Serious+Fatal)
Dependent variable: Injury severity (class 1 = Unharmed+Minor, class 2 = Serious+Fatal)
Independent variables: Temporal (Hour H, Day of the week DW, Month M), Road environment (Luminosity L, Weather W), Road characteristics (Horizontal design HD, Vertical design VD, Shoulder S, Circulation direction CD, Intersection I, Vertical signs VS, Horizontal signs HS, Light signs LS, Pavement type PT, Pavement grip PG, Pavement condition PC, Local speed LSp), Pedestrian (Age A, Sex, Type T, Pedestrian action PA)
*All cases of the dependent variable classified in class 1

with all variables classes are presented and analyzed. Although the percent of incorrect predictions obtained for the training and testing phases is acceptable (7.1% to 9.9%), the classification results show that all cases are classified in class 1 (unharmed or minor injured pedestrians). To evaluate the adequacy of the modeling process options, several analyzes were carried out considering SPSS available options, however, similar results were obtained. The model outcomes can be justified by the significant imbalance between the dependent variable classes and possibly because accidents involving seriously injured and fatal victims do not present significantly different characteristics from accidents involving unharmed and minor injured. To confirm the previous statement and evaluate the discriminant power of the models, ROC curves were produced for each dependent variable category. All curves presented an area under the curve between 0.5 and 0.7, indicating a weak discrimination ability [20].

Concerning the risk factors, 7 of the variables identified in the 'Top10 independent variable importance' of the two complete models were the same: LSp: 51–90 km/h, L: Night, no lighting, LS: Off, S: Unpaved shoulder, PT: Cement, PC: Poor and T: Moving on skate, scooter, or others. For the road characteristics models, 7 of the most important variables were also the same (3 identified in the complete models): LSp: 0–30 km/h, LSp: 51–90 km/h, S: Unpaved shoulder, PT: Stone, PC: Poor, VS: Stop and LS: Intermittent. A similar situation occurs with pedestrian models, with the 7 common variables: A: 0–9 years old, A: 10–17 years old, T: Isolated pedestrian, T: Pedestrian in group, T: Moving on skate, scooter or others, T: Pushing a bicycle, stroller or people with physical disabilities and PA: Crossing a marked crossing without respecting traffic lights. For the pedestrian models, it appears that the age classes 50–65 and > 65 years old, identified in the pre-Covid19 period, are replaced by the young and middle age classes (18–24 and 25–49 years old) in the Covid19 period, highlighting the increased care with seniors in the pandemic period. From the results analysis it is possible to suggest interventions aimed at improving the lighting conditions and traffic lights operation in urban pedestrian crossings, an adequate maintenance of shoulders and pavements, especial attention to urban areas with higher speed limits (51–90 km/h) and raising awareness among pedestrians moving in soft modes or pushing strollers or people with physical disabilities.

## 3 Conclusions

As main conclusions, it is possible to highlight that, although the prediction models produced for pedestrian injury severity present a high percentage of correct classifications, the evidence that pedestrian injuries are essentially minor (around 92%) and possible little differentiation between factors contributing to 'serious injuries+fatalities' and 'unharmed+minor injuries', may justify the weak discriminating power of the models. Regarding the risk factors, the results obtained revealed a certain consistency between models, suggesting that the main interventions aimed at reducing pedestrian casualties should focus on improving pedestrian crossings traffic lights operation, lighting, shoulders and pavement conditions. Special attention should also be paid to urban areas with higher speed limits (51–90 km/h) and in raising awareness among pedestrians moving in soft modes or pushing strollers or wheelchairs. Regarding the comparison between the pre-Covid19 and Covid19 scenarios, a significant decrease was verified in the annual number of accidents (around 30–35%) and differences were found for some pedestrian age classes. The importance of classes 18–24 and 25–49 years old increased in the Covid19 period contrasting to the previous importance of 50–64 and >65 years old in the pre-Covid 19 period. Future developments include evaluating other software and machine learning algorithms to identify those that can better accommodate categorical variables and less balanced distributions of the dependent variable classes.

**Acknowledgement.** The authors are very grateful for the support granted by the Research Unit GeoBioTec, through the project reference UIDB/04035/2020, funded by the Fundação para a Ciência e a Tecnologia, IP/MCTES through national funds (PIDDAC).

## References

1. Safety, P.: Urban space and health. OECD, In ITF Research Reports (2012)
2. Meocci, M., Branzi, V., Martini, G., Arrighi, R., Petrizzo, I.: A predictive pedestrian crash model based on artificial intelligence techniques. Appl. Sci. **11**(23), 11364 (2021)
3. Eurostat. https://ec.europa.eu/eurostat. Accessed 9 Oct 2023. Persons killed in road accidents by sex (source: CARE)
4. Autoridade Nacional de Segurança Rodoviária: Vitimas Mortais a 30 Dias (2010)
5. Autoridade Nacional de Segurança Rodoviária: Relatório Anual 2020 - Sinistralidade 30 dias (2022)
6. European Road Safety Observatory: Facts and Figures - Pedestrians 2023 (2023)
7. European Road Safety Observatory: Facts and Figures - Urban areas – 2022 (2022)
8. Kamboozia, N., Ameri, M., Hosseinian, S.: Statistical analysis and accident prediction models leading to pedestrian injuries and deaths on rural roads in Iran. Int. J. Inj. Contr. Saf. Promot. **27**(4), 493–509 (2020)
9. Ali, Y., Haque, M., Mannering, F.: A Bayesian generalised extreme value model to estimate real-time pedestrian crash risks at signalised intersections using artificial intelligence-based video analytics. Anal. Methods Accid. Res. **38**, 100264 (2023)
10. Zhang, S., Abdel-Aty, M., Cai, Q., Li, P., Ugan, J.: Prediction of pedestrian-vehicle conflicts at signalized intersections based on long short-term memory neural network. Accid. Anal. Prev. **148**, 105799 (2020)
11. Kuşkapan, E., Sahraei, M., Çodur, M.K., Çodur, M.Y.: Pedestrian safety at signalized intersections Spatial and machine learning approaches. J. Transp. Health **24**, 101322 (2022)
12. Chakraborty, A., Mukherjee, D., Mitra, S.: Development of pedestrian crash prediction model for a developing country using artificial neural network. Int. J. Inj. Contr. Saf. Promot. **26**(3), 283–293 (2019)
13. Hossain, A., Sun, X., Zafri, N., Codjoe, J.: Investigating pedestrian crash patterns at high-speed intersection and road segments: findings from the unsupervised learning algorithm. Int. J. Transport. Sci. Technol. (2023)
14. Noh, B., Yeo, H.: A novel method of predictive collision risk area estimation for proactive pedestrian accident prevention system in urban surveillance infrastructure. Transp. Res. Part C Emerg. Technol. **137**, 103570 (2022)
15. Olowosegun, A., Babajide, N., Akintola, A., Fountas, G.: Analysis of pedestrian accident injury-severities at road junctions and crossings using an advanced random parameter modelling framework: the case of Scotland. Accid. Anal. Prev. **169**, 106610 (2022)
16. Mirhashemi, A., Amirifar, S., Kashani, A., Zou, X.: Macro-level literature analysis on pedestrian safety: bibliometric overview, conceptual frames, and trends. Accid. Anal. Prev. **174**, 106720 (2022)
17. Shrinivas, V., Bastien, C., Davies, H., Daneshkhah, A., Hardwicke, J.: Parameters influencing pedestrian injury and severity - a systematic review and meta-analysis. Transport. Eng. **11**, 100158 (2023)
18. Silva, P., Andrade, M., Ferreira, S.: Machine learning applied to road safety modeling. J. Traffic Transport. Eng. **7**(6), 775–790 (2020)

19. Mokhtarimousavi, S., Anderson, J., Azizinamini, A., Hadi, M.: Factors affecting injury severity in vehicle-pedestrian crashes: a day-of-week analysis using random parameter ordered response models and Artificial Neural Networks. Int. J. Transport. Sci. Technol. **9**(2), 100–115 (2020)
20. Hosmer, D., Lemeshow, S., Sturdivant, R.: Applied logistic regression, 3rd edn. Wiley, New Jersey (2013)