Pedestrian Crash Severity Prediction and Contributory Factors Analysis by Using Machine Learning Methods

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

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

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

Abstract. Pedestrians occupy a leading position among the most vulnerable road users. Each year about 270,000 pedestrians die due to road accidents, so this study aims to highlight the most influencing contributory factors and the most promising models to predict pedestrian crash severity. ISTAT data for the City of Rome (2013–2020) are used and different Machine Learning Methods are trained and tested, after balancing the data with oversampling techniques. In addition, analysis of the most influencing contributory factor is carried out, by using the ROC curve method, Variable Importance Analysis (VIP), and Support Vector Machine with a Linear Kernel. The findings suggest that the model with the best prediction performance is the Random Forest, followed by the Decision Tree and k-nearest neighbour algorithm. Regarding the analysis of contributory factors, the methods implemented highlight that the hour in which the accident occurs, pedestrian gender, and age seem to be the most critical factors that increase the severity of a pedestrian crash. There are also some limitations in this study: the first is connected to the black-box nature of these models; the second regards how these variables could influence positively or negatively the outcome.

Keywords: Sustainable Mobility · Road Safety · Pedestrian accidents · pedestrian safety · Random Forest · Support Vector Machines (SVM) · Machine Learning

1 Introduction

Road accidents are among the top ten causes of death worldwide, ranking seventh overall. On the other hand, they occupy the first place as regards deaths between the ages of 15 and 29 [1]. More than 1.2 million people in a single year lose their lives and most of them incur not-fatal injuries that lead often to injury [2].

Although the damages to the human body are unfortunately well known, thanks also to the current scientific literature review [3–7], accidents may cause a large economic

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
O. Gervasi et al. (Eds.): ICCSA 2025, LNCS 15898, pp. 3–14, 2026.
https://doi.org/10.1007/978-3-031-97657-5_1

toll [8], and knowing and understanding this macroeconomic burden could give a strong direction in an effective policy-making decision process.

Among the most vulnerable road users, pedestrians occupy a leading position. Pedestrian accidents are a serious public health issue that can cause injuries, disabilities, and even death. According to the World Health Organization, more than 1.3 million people die each year as a result of road traffic crashes, and about 20% (270,000) of them are pedestrians [9, 10].

Pedestrian accidents can be caused by various factors, such as human error, vehicle design, road environment, weather conditions, and enforcement of traffic rules. Some of the common causes of pedestrian accidents are speeding, lack of visibility, alcohol, distractions, and road design.

Higher vehicle speeds increase both the likelihood of a pedestrian being struck by a car and the injury severity [11]. Most pedestrian deaths (60% in 2020) occur on high-capacity urban roads that typically have posted speed limits of 45–55 miles per hour [12].

Poor lighting, especially at night or in bad weather, can make it difficult for pedestrians to see approaching vehicles or hazards. Some authors suggest that pedestrians should wear reflective clothing or use flashlights when walking in low-light conditions [13].

Driving under the influence of alcohol or other psychoactive substances impairs a driver's ability to react to changing situations and increases the risk of crashing into pedestrians. Alcohol is involved in about 30% of all road traffic fatalities worldwide [13].

Using a mobile phone or other devices while walking can divert a pedestrian's attention from the road and increase the chance of collision, so the authors suggest that pedestrians should avoid using their phones or other devices while walking and keep them out of sight [14].

Streets designed to be narrower and for slower speeds experience the lowest rates of vehicle-pedestrian crashes, and those with wide lanes and higher operating speeds experience the highest rates [15]. Effective countermeasures include sidewalks, pedestrian-only signal phasing, and optimal roadway lighting [16].

According to the literature [9, 17], changes to the road environment and legislation and a higher focus on education effectively reduce pedestrian accidents.

In this paper, several Machine Learning Methods are used to classify pedestrian accidents. The dependent variable is the outcome of the crash and so the severity, i.e., if the road accident led to an injury or death. The study aims also to discover the role of the main variables in pedestrian accidents, by using Machine Learning Methods.

2 Crash Data

To choose the most adequate dataset, it was necessary to review a series of data collections over several years, focusing more attention on accidents involving vulnerable users (namely, cyclists and pedestrians). The area in which this analysis is performed is the City of Rome. The main data that could be used are those provided by ISTAT (National Institute of Statistics) [18] and by the City of Rome through "Rome mobility" which,

despite being similar in content, have some differences [19]. Finally, the ISTAT dataset was chosen because it covers more years than the other one.

ISTAT is a public body that carries out numerous researches, surveys, and censuses in Italy, providing essential information for statistical research and for numerous fields of study. ISTAT records, using checks and police reports, incidents in which at least one person involved is injured or dead, with a maximum viewing time of 30 days, further classifying the dead as deceased within 24 h or 30 days.

At the national level, the data are recorded thanks to the collaboration of the police in charge of intervening at the site of the accident. ISTAT sets the annual dataset by recording some key features:

- Date
- Time
- Accident Location
- Road type
- Road surface
- Signposting information
- Weather conditions
- Nature of the accident
- Type of vehicles involved
- Consequences of the accident
- Personal data of drivers, passengers and pedestrians
- Circumstances of the accident

3 Method

Machine Learning (ML) is an Artificial Intelligence (AI) branch. It is renewed as an important tool to analyse large amounts of data and use models to forecast the dependent variable, in the context of regression or classification tasks [20]. With Machine Learning it is possible to develop algorithms that can analyse the data provided and their properties and characteristics. Another important feature of such types of models is that they can learn and become more and more stable and reliable by putting inside more and more data [20, 21].

In Fig. 1, the flow chart of the common methodology that is used with Machine Learning Methods is here proposed. The first step is the data collection, that is needed to train and test the several models. Different data sources could be used and this depends on the type of field or task that needs to be undertaken. After this collection phase, datasets were cleaned, pre-processed, and transformed or standardized properly to complete the classification task. Then, several models proposed according to the most used models in the scientific literature are trained on the training set (usually 70% of data in the data set are used) and finally, the performance metrics for each model are evaluated on the test set. The best model is selected according to the best prediction performances. The reason why focusing on Machine Learning Methods is that they show better performances according to the current literature review.

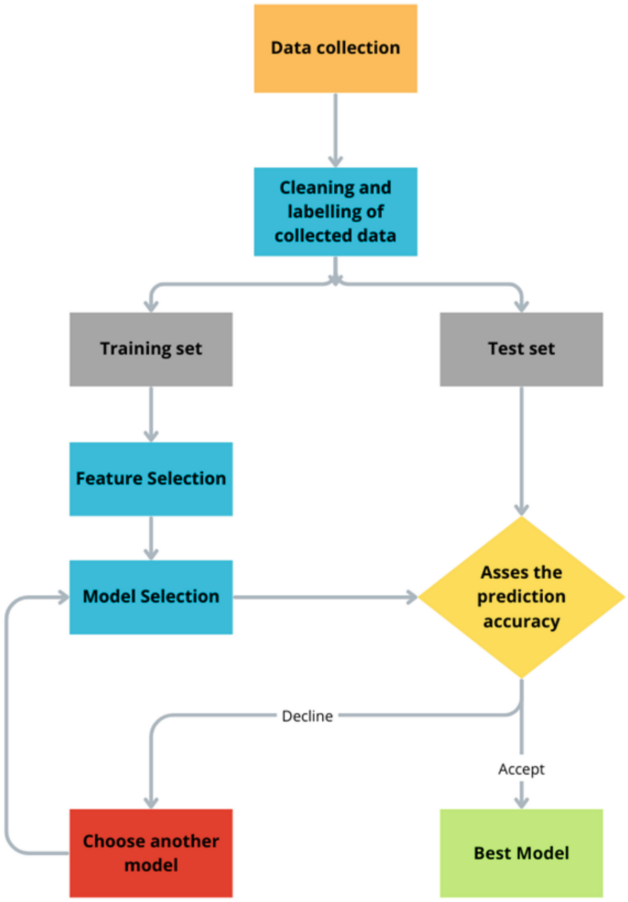

Fig. 1. Flow chart of a common Machine Learning Process

3.1 Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA)

Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) are two common classifiers, among the Discriminant Analysis (DA) methods [22]. As the names suggest, the decision boundary in LDA is a linear plan [23], while in QDA is a non-linear one. The models aim to find the best hyperplane that maximizes the distance between the several observations. The two methods are often used as a classifier but sometimes also as a feature selection method [24].

3.2 K-Nearest Neighbour (KNN)

K-Nearest Neighbour (KNN) is a classifier first introduced by Fix and Hodges [25]. The method is based on the assumption that similar and related features create a different cluster in the whole space of features [26]. That means that an object in this space is

classified under a class or another based on the characteristics of the nearest neighbours. The number of neighbour is represented by the parameter k: if the magnitude of k increases, the complexity of the model increases, while the errors could decrease. For this reason, the number of neighbours is always an even number, to avoid uncertainties in the classification.

3.3 Decision Tree (DT) and Random Forest (RF)

Decision Tree (DT) and Random Forest (RF) are two Machine Learning methods used for both classification and regression. A DT works by creating some partition within the original data in sub-groups [27]. These regions are also called leaves that are connected with branches to internal nodes that define a hierarchical structure. The structure of a tree is made hierarchically, by splitting the feature space according to some measure of homogeneity [28]. To avoid a high number of splits, the tree is pruned. Decision Tree Models are characterized by high variance and to avoid this problem some other algorithms, such as Random Forest, are introduced. The main idea behind RF is to collect information from several Decision Trees and then combine them to improve prediction accuracy [29–31].

3.4 Support Vector Machine (SVM)

The development of Support Vector Machines (SVM) dates back to the 1990s. SVM aims to find a hyperplane that separates observations to make classifications [32]. To find the best hyperplane, it is possible to maximize the margin (the distance between the plane and the closest observation). An important parameter used for the optimization phase is the tuning parameter cost (c): when the value of this parameter is higher, the wider the margin and the more stable will be the hyperplane [33]. The feature space could be also enlarged and so transformed into a higher-dimensional feature space with a kernel function, to handle the non-linear relationship between dependent and independent variables. In this new transformation of the space, the decision boundary is a linear function, but in the original feature space could be, for instance, a radial or a polynomial boundary [26].

3.5 Variable Importance Analysis (VIP)

To complete an analysis of the most influencing factor, a Variable Importance Analysis (VIP) could be implemented. VIP is a built-in algorithm in the Random Forest method: the importance of a variable is evaluated as a mean decrease in accuracy by using out-of-bag observations (OOB) ($L_{b,oob}$), because the tree grows from a bootstrapped sample (L_b) from the original training sample (L) [34].

Another method to select classifiers is the Receiver Operating Characteristic Curve (ROC) graph tool [35, 36], usually used instead of an accuracy performance measure that in some cases could not be an adequate metric [37, 38].

4 Results

In this section, the main results of the analysis carried out are shown. First, comparisons between the Machine Learning Algorithms are proposed, by evaluating the F1 score of the different models. The F1 score is a common metric that is often used for these purposes [21]. In addition to that, an analysis of contributory factors is performed to cover the gap between the lack of interpretability that is often attributed to Machine Learning Methods.

4.1 Model Performances

F measures, such as the F1-score, represent a common way to assess Machine Learning Method performances. These metrics give information on how the model is reliable and accurate. To define these metrics, it is essential to introduce the confusion matrix concept. A confusion matrix is a sort of way to visualize the results of classification methods and how many observations have been classified in the correct or wrong way by the Machine Learning Algorithm. In the confusion matrix, the number of True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN) are organized as it is possible to see in Fig. 2.

	Actual Values	
Predicted Values	True Positive (TP)	False Positive (FP)
	False Negative (FN)	True Negative (TN)

Fig. 2. Example of a confusion matrix.

The four elements of the matrix, on the diagonal and the anti-diagonal, describe the relationships between the actual values (that represent the real observation in the dataset) and the model's forecast. It is so possible to summarize:

- The True Positives (TP) are the observation that the forecasting model has correctly classified as a positive class and that in reality, they belong to the positive class (for instance, the model predicts some accidents with an injury, and in fact, the accident led to an injury);
- The True Negatives (TN) are the observation that the forecasting model has correctly classified as a negative class and that in reality, they belong to the negative class (for instance, the model predicts some accidents with a death, and in fact, the accident led to a death);
- The False Positives (FP) are the observation that the forecasting model has correctly classified as a positive class and that in reality, they belong to the negative class (for

instance, the model predicts some accidents with an injury, and in fact, the accident led to a death). The error related to this misclassification is usually known as Type I error;

- The False Negatives (FN) are the observation that the forecasting model has correctly classified as a negative class and that in reality, they belong to the positive class (for instance, the model predicts some accidents with death, and in fact, the accident led to an injury). The error related to this misclassification is usually known as Type II error.

The F1 score in Eq. 1 (that is used as the main metric to assess performances and compare them) could be considered as the harmonic mean of recall (Eq. 2) and precision (Eq. 3):

$$\mathrm{F1-score} = \frac{2 * \mathrm{Precision} * \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \tag{1}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{2}$$

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{3}$$

In Table 1, the performances in terms of F1 score are listed for each model analyzed in this paper, and they are compared with a prior study [21]. As highlighted, in this paper the performances of QDA and kNN are lower than the previous models, Support vector machines with different kernels remain quite the same, but there is a huge improvement with decision tree and random forest, compared to previous models.

Table 1. Performance metrics of the analyzed models.

Method	Current study	Prior study [21]
	F1-score	F1-score
Linear Discriminant Analysis (LDA)	0.6582	0.6741
Quadratic Discriminant Analysis (QDA)	0.7909	0.7594
K-Nearest Neighbor (KNN, k = 3)	0.9007	0.9370
K-Nearest Neighbor (KNN, k = 5)	0.9011	–
K-Nearest Neighbor (KNN, k = 7)	0.8976	–
Decision Tree (DT)	0.9821	0.9639
Random Forest (RF)	0.9824	0.9678
Support Vector Machine with Linear Kernel (SVM-L)	0.6922	0.6367
Support Vector Machine with Radial Kernel (SVM-P)	0.7674	0.7687
Support Vector Machine with Polynomial Kernel (SVM-R)	0.8771	0.8764

4.2 Variable Importance Analysis

As further results of this analysis, a variable importance analysis is carried out to highlight the main contributory factors in pedestrian road crashes. As it is possible to see in Fig. 3, there are some Machine Learning Methods that could be used to perform such a type of analysis. For the ROC curve method, the algorithm evaluates and calculates the area



Fig. 3. Comparison of the selected feature Selection algorithm and contributing factors in pedestrian road crashes.

under the ROC curve when a single independent variable is compared with the dependent variable: the higher this area, the higher the correlation between the dependent and the independent variable. Random Forest and Decision Tree have a built-in algorithm to evaluate the most influential features (as is possible to read in the section "Methods"). SVM with linear kernel does not have such a type of feature importance built-in the algorithm: for this reason, to perform an analysis on contributing factors, it is possible to have an idea of which variable is highly related to the outcome by comparing the estimate of the coefficients of the hyperplane that divides observations.

Among the top selected features, pedestrian-related features are the most important ones. Pedestrian age is one of the most influential factors according to the ROC curve method, Random Forest, and Decision Tree. Only in the SVM model is it in 8th place. Pedestrian Gender occupies the top five in all selected methods, highlighting that males and females, in addition to their age, resist crashes based on their different vulnerability. Another variable is the hour of the day, and this is due mainly to the different traffic road conditions in terms of vehicular flows and speeds. Other selected variables are Pedestrian Behaviours and Stretch (intersection or road, in the ROC curve Method), Pavement (Decision Tree and Random Forest), Day (Decision Tree), Location (Random Forest), Surface, Vehicle Type and Signposting (SVM).

5 Conclusions

A review of current scientific and technical literature highlights the need to identify the main contributory factors that contribute to accidents. Determining effective measures to mitigate the severity of injuries among vulnerable road users is a leading priority. Given the complexity of this issue, Machine Learning Models are often used. In this study, an initial attempt was made to predict accident outcomes using ISTAT data from Rome's metropolitan area. The research began with the training and testing of multiple Machine Learning Models. Model performance was evaluated on the test set to ensure validity. Several models were assessed based on their complexity, with no bias toward any particular approach. The Random Forest model demonstrated the highest F1 score, and so the highest capabilities to predict the outcome of a pedestrian crash.

The results of the analysis of contributory factors (see Fig. 3) show that a main role is played by the hour, pedestrian age, and pedestrian gender: all four proposed methods (in particular the ROC curve Method, Decision Tree, Random Forest, and Support Vector Machine with Linear kernel) select these three variables among the first five positions of the ranking (exception made for SVM, where pedestrian age is on the 7th place). The hour, such as the pavement and the location (as selected from Decision Tree and Random Forest) could give an idea for which reason they are so important: the hour is something related to traffic flows, while pavement and location to vehicle speeds.

Despite these results, the study presents some limitations. The first is connected to the black-box nature of these models because there is some lack on the interpretability side. The second regards the methods of variable importance analysis. As it is possible to see in Fig. 3, the output of these analyses is only a ranking of the most influential features that contribute to predicting pedestrian accident severity, and not how these could positively or negatively affect the outcome. For instance, if the variable "Hour" is

selected, it is difficult to understand how daylight or nighttime hours increase or decrease the probability of fatal crashes. Another limitation is that the analysis is carried out only in the city of Rome, so the model is well trained and tested for this context. From this last sentence, as further steps, it could be interesting if the same variables selected with the aforementioned methods are the most influential also in other cities in Italy. Another future development aims to combine models and different methodologies to improve the interpretability of the models and understand how a single variable quantitatively affects the pedestrian severity. A point in favour of this methodology and models is that they could be replicated in different contexts and also on different scales, from the city level to the national level. This is also a first step in pedestrian accident analysis, but the findings could be useful to be a reference point to improve safety conditions for pedestrians in cities. The analysis carried out shows that pedestrian-related variables are the most influential factors, alongside variables that could be related to traffic conditions.

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

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

References

1. WHO: Global status report on road safety 2015. World Health Organization, Geneva (2015)
2. World Health Organization (WHO): Road traffic injuries
3. Shang, S., et al.: The predictive capacity of the MADYMO ellipsoid pedestrian model for pedestrian ground contact kinematics and injury evaluation. Accid. Anal. Prev. **149**, 105803 (2021)
4. McNally, D.S., Rosenberg, N.M.: MADYMO simulation of children in cycle accidents: a novel approach in risk assessment. Accid. Anal. Prev. **59**, 469–478 (2013)
5. Guo, R., Yuan, Q., Sturgess, C.E.N., Hassan, A.M., Li, Y., Hu, Y.: A study of an Asian anthropometric pedestrian in vehicle–pedestrian accidents using real-world accident data. Int. J. Crashworth. **11**(6), 541–551 (2006)
6. Mizuno, K., et al.: Analysis of fall kinematics and injury risks in ground impact in car-pedestrian collisions using impulse. Accid. Anal. Prev. **176**, 106793 (2022)
7. Nie, J., Li, G., Yang, J.: A study of fatality risk and head dynamic response of cyclist and pedestrian based on passenger car accident data analysis and simulations. Traffic Inj. Prev. **16**(1), 76–83 (2015)
8. Chen, S., Kuhn, M., Prettner, K., Bloom, D.E.: The global macroeconomic burden of road injuries: estimates and projections for 166 countries. Lancet Planet. Health **3**(9), e390–e398 (2019)

9. Namatovu, S., et al.: Interventions to reduce pedestrian road traffic injuries: a systematic review of randomized controlled trials, cluster randomized controlled trials, interrupted time-series, and controlled before-after studies. PLoS ONE **17**(1), e0262681 (2022)
10. World Health Organization: Pedestrian safety: a road safety manual for decision-makers and practitioners. World Health Organization (2023)
11. Road Safety (WHO). https://www.who.int/health-topics/road-safety#tab=tab_2
12. Kendi, S., Johnston, B.D.: Epidemiology and prevention of child pedestrian injury. Pediatrics **152**(1), e2023062508 (2023)
13. Pedestrian Safety: Prevent Pedestrian Crashes. https://www.nhtsa.gov/road-safety/pedestrian-safety
14. European Road Safety Observatory. https://road-safety.transport.ec.europa.eu/system/files/2021-07/facts_figures_pedestrians_final_20210323.pdf
15. Road Safety Statistics. https://transport.ec.europa.eu/background/2021-road-safety-statistics-what-behind-figures_en
16. Road Traffic Injuries. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries
17. Fisa, R., Musukuma, M., Sampa, M., Musonda, P., Young, T.: Effects of interventions for preventing road traffic crashes: an overview of systematic reviews. BMC Public Health **22**(1), 513 (2022)
18. Istat. https://www.istat.it/. Accessed July 2024
19. Roma Mobilità. https://romamobilita.it/it. Accessed Oct 2024
20. Pugliese, R., Regondi, S., Marini, R.: Machine learning-based approach: global trends, research directions, and regulatory standpoints. Data Sci. Manag. **4**, 19–29 (2021)
21. D'Apuzzo, M., Cappelli, G., Naranni, S., De Guidi, M., Nicolosi, V.: Machine learning tools for predicting the outcome of pedestrian crashes: preliminary findings in the Metropolitan city of Rome. In: Gervasi, O., Murgante, B., Garau, C., Taniar, D., C. Rocha, A.M.A., Faginas Lago, M.N. (eds.) ICCSA 2024. LNCS, vol. 14824, pp. 116–132. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-65332-2_8
22. Mika, S., Ratsch, G., Weston, J., Scholkopf, B., Mullers, K.R.: Fisher discriminant analysis with kernels. In: Neural Networks for Signal Processing IX: Proceedings of the 1999 IEEE Signal Processing Society Workshop (Cat. No. 98TH8468), pp. 41–48. IEEE, August 1999
23. Tharwat, A.: Linear vs. quadratic discriminant analysis classifier: a tutorial. Int. J. Appl. Pattern Recogn. **3**(2), 145–180 (2016)
24. Makaba, T., Doorsamy, W., Paul, B.S.: Exploratory framework for analysing road traffic accident data with validation on Gauteng province data. Cogent Eng. **7**(1), 1834659 (2020)
25. Fix, E., Hodges, J.: Discriminatory analysis. Nonparametric discrimination: Consistency properties. Technical Report 4, US AF School of Aviation Medicine, Randolph Field, Texas (1951)
26. Sarraf, J., Pattnaik, P.K.: Brain-computer interfaces and their applications. In: An Industrial IoT Approach for Pharmaceutical Industry Growth, pp. 31–54. Academic Press (2020)
27. James, G., Witten, D., Hastie, T., Tibshirani, R.: An Introduction to Statistical Learning, vol. 112, p. 138. Springer, New York (2013)
28. Chen, M.M., Chen, M.C.: Modeling road accident severity with comparisons of logistic regression, decision tree and random forest. Information **11**(5), 270 (2020)
29. Elyassami, S., Hamid, Y., Habuza, T.: Road crashes analysis and prediction using gradient boosted and random forest trees. In: 2020 6th IEEE Congress on Information Science and Technology (CiSt), pp. 520–525. IEEE, June 2021
30. Rella Riccardi, M., Galante, F., Scarano, A., Montella, A.: Econometric and machine learning methods to identify pedestrian crash patterns. Sustainability **14**(22), 15471 (2022)
31. Scarano, A., Riccardi, M.R., Mauriello, F., D'Agostino, C., Pasquino, N., Montella, A.: Injury severity prediction of cyclist crashes using random forests and random parameters logit models. Accid. Anal. Prev. **192**, 107275 (2023)

32. Cesarini, L., Figueiredo, R., Monteleone, B., Martina, M.L.: The potential of machine learning for weather index insurance. Nat. Hazards Earth Syst. Sci. **21**(8), 2379–2405 (2021)
33. Mas, J.F., Flores, J.J.: The application of artificial neural networks to the analysis of remotely sensed data. Int. J. Remote Sens. **29**, 617–663 (2008). https://doi.org/10.1080/01431160701352154
34. Archer, K.J., Kimes, R.V.: Empirical characterization of random forest variable importance measures. Comput. Stat. Data Anal. **52**(4), 2249–2260 (2008)
35. Fawcett, T.: An introduction to ROC analysis. Pattern Recogn. Lett. **27**, 861–874 (2006)
36. Güvenir, H.A., Kurtcephe, M.: Ranking instances by maximizing the area under ROC curve. IEEE Trans. Knowl. Data Eng. **25**(10), 2356–2366 (2012)
37. Provost, F., Fawcett, T.: Analysis and visualization of classifier performance: comparison under imprecise class and cost distributions. In: Proceedings of the Third International Conference on Knowledge Discovery and Data Mining, pp. 43–48 (1997)
38. Provost, F., Fawcett, T., Kohavi, R.: The case against accuracy estimation for comparing induction algorithms. In: Proceedings of the 15th International Conference on Machine Learning, pp. 445–453 (1998)