From Trend Observation to Risk Interpretation: A Machine Learning Analysis of Micromobility Injury Severity in 2022

Paraskevi Koliou[1*[0000-0003-3926-3345]], Eleni Maria Theodoraki[1[0009-0001-6033-0728]], George Yannis[1[0000-0002-7017-0756]]

[1] National and Technical University of Athens, Iroon Polytechniou 5-9, 15772 Athens, Greece
*paraskevikoliou@gmail.com

Abstract. Longitudinal analysis of micromobility-related traffic collisions across Europe (2013–2023) shows a steady increase, with 2022 marking a pronounced spike across all severity categories. This study conducts a severity-risk assessment for 2022, identifying key factors influencing serious injury outcomes. Using multi-country accident data, we trained Random Forest models distinguishing between micromobility users and other vehicles. Variables included demographics, environmental conditions, lighting, and spatial context. Interpretability was ensured with SHapley Additive exPlanations (SHAP). Results show that for micromobility users, age is the strongest predictor, with heightened risk among younger (18–24) and older (65+) riders. Weather, lighting, and rural settings also contributed. For other users, rural context, weather, and male gender dominated. Models achieved moderate predictive performance (63.1% and 67.2%), with SHAP providing clear insights. This approach highlights differentiated risk profiles and supports targeted, evidence-based safety interventions.

Keywords: Micromobility, Road Safety, SHAP, Machine Learning, Random Forest

1 Introduction

Micromobility has grown rapidly, becoming an integral part of urban transportation as it offers environmentally friendly, low-emission modes of transport for short distances. Micromobility users, unprotected by any vehicle body, tend to be more physically exposed, and they mostly share space with vehicles in mixed-traffic environments where the provision of suitable infrastructure is still uneven [1]. The common use of the same road and the scarce presence of protective elements further increase the risk of accidents, especially in places where infrastructure is lacking or is in poor condition [2].

Therefore, the pattern of injuries resulting from these incidents is structurally different from that of typical road users.

Researchers have found several factors that determine the severity of injuries sustained by micromobility users. Vulnerability profiles of individuals greatly depend on demographics, such as age and gender [3]. Environmental and behavioral factors like the use of the helmet, speed, the discipline of the lane, and the characteristics of the trip also influence the results of the injury [4,5,6]. Moreover, features of the built environment strongly affect the nature of the accident and the degree of injury. Studies have demonstrated that urban typology, commercial land use, and the spatial patterns of clustering are significant determinants of the severity of accidents [7,8]. This evidence indicates that the risk of injury is highly dependent on the situation and not evenly spread over the area. Random Forest, AdaBoost, and neural networks are some of the algorithms that have shown better predictive capabilities than conventional regression models [5,6]. Further improvements in explanatory power have been achieved with hybrid systems that combine clustering, spatial analysis, and statistical modeling [7,8]. Nevertheless, many ML models are often criticized for their limited interpretability. To solve this problem, there have been advances in explainable AI (XAI) methods such as SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME), which can explain how much each factor contributes to the final prediction [9,10]. Some very recent uses of SHAP show how the method can uncover the effects of infrastructure, user characteristics, and contextual factors on the outcomes of micromobility [10].

Despite growing interest, few studies integrate interpretable ML with harmonized multi-country crash datasets or directly compare micromobility and other road users within a unified framework. This study addresses these gaps by combining longitudinal descriptive analysis and interpretable ML using a standardized European crash database (2013–2023), focusing on injury severity in 2022. The objective is to identify differentiated severity mechanisms and support targeted, evidence-based road safety interventions.

2 Methodology

2.1 Data Source and Preparation

Standardized traffic accident data from a multi-country European database covering the years 2013–2023 were used in this study. Only cases with full injury severity and vehicle type information were included in the harmonized data, which was compiled from yearly Excel files. Micromobility users were defined as pedal cycle and moped riders in accordance with official crash reporting standards. Mopeds were included due to their comparable operational characteristics (low mass, limited speed, urban usage).

Severity was originally recorded as fatal, serious, and slight. For modelling, a binary classification (serious/fatal vs. slight) was adopted to maintain class balance and enhance interpretability. Although predictive modelling focused on 2022, longitudinal trends were analyzed to contextualize this year.

2.2 Descriptive Statistical Analysis

The annual descriptive statistics for micromobility injuries between 2013 and 2023 are provided in the Supplementary Material. Overall, the results indicate a steady increase in total injuries over the study period, with slight injuries consistently representing the largest share. Serious injuries showed greater variability but a gradual upward trend, while fatal injuries remained relatively stable. A temporary decline in 2020 is consistent with mobility restrictions during the COVID-19 pandemic, followed by a pronounced rebound in 2022. These longitudinal patterns confirm that the selected modelling year reflects broader structural trends rather than an isolated anomaly.

A closer look at 2022 revealed important contrasts between micromobility and other road users. Counts represent absolute injury frequencies per severity level. As shown in Table 1, micromobility riders experienced a lower proportion of fatal injuries (3.7% vs. 6.4%) but higher proportions of serious and slight injuries, and significantly higher rates of serious and slight injuries, suggesting heterogeneous exposure and supporting the need for mode-specific modelling.

Table 1. 2022 Descriptive Statistics by Mode Type

	Fatally Injured (at 30 days)		
	count	**mean**	**std**
Other Vehicles	264893	0.064	0.28
Micromobility	63859	0.037	0.20
	Seriously Injured (as reported)		
	count	**mean**	**std**
Other Vehicles	264893	0.31	0.87
Micromobility	63859	0.50	1.41
	Slightly Injured		
	count	**mean**	**std**
Other Vehicles	264893	2.06	4.88
Micromobility	63859	2.46	7.74

These patterns motivated the development of mode-specific models to capture better differences in crash severity mechanisms between micromobility and other users.

2.3 Model Development, Evaluation and Interpretation

To identify varying severity mechanisms in micromobility and non-micromobility users, two separate Random Forest classifiers were trained for each group. The target variable was a binary outcome (serious/fatal versus slight injury), a formulation that was determined to bring about a better class balance and thus facilitate interpretability. Hyperparameters (n_estimators, max_depth, and min_samples_leaf) were optimized using grid search with stratified 5-fold cross-validation, and model performance was reported as mean and standard deviation across folds. To tackle the problem of class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was administered solely on the training folds (k = 5; 1:1 class ratio), thus data leakage was prevented. Missing data were imputed by taking the median for continuous

Model evaluation prioritized accuracy, precision, recall, and F1-score, reflecting policy-relevant error patterns rather than purely predictive metrics. To make the model more comprehensible, the results were analyzed with SHapley Additive exPlanations (SHAP), which provides the means to quantify both the size and the sign of the contribution of a feature towards the predicted injury severity.

3 Results

3.1 Descriptive Trends

Between 2013 and 2023, micromobility-related injuries increased steadily across Europe, with slight injuries accounting for most cases. Fatalities remained relatively stable at fewer than 3,000 annually, while serious injuries showed greater variability but an overall upward trend. A decline in 2020 coincided with COVID-19 mobility restrictions, followed by a pronounced rebound in 2022 (Figure 1).

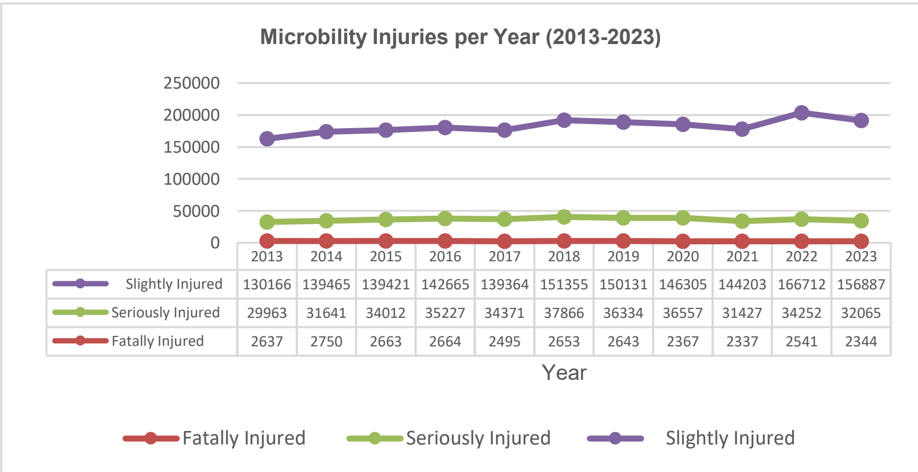

Fig. 1. Micromobility Injuries per Year

Year-on-year changes indicate increased volatility after 2018, likely reflect changing travel behaviour and post-pandemic recovery. Riders aged 25–49 accounted for the largest number of injuries, while younger (18–24) and older (65+) users showed elevated severity risk. This potential non-linear age pattern should be interpreted cautiously, as it reflects marginal SHAP effects rather than formal statistical testing. Male riders consistently exhibited higher involvement across working-age groups. Although most crashes occurred under clear or unreported weather conditions, adverse weather was associated with higher severity. Similarly, crashes were more frequent during daytime and weekdays, but severe outcomes were more likely to reduce visibility and lower traffic volumes. Spatial patterns further highlighted contextual differences, with rural environments associated with a higher likelihood of severe and fatal outcomes.

3.2 Model Performance

The micromobility model achieved an accuracy of 63.1% (± SD), while the non-micromobility model achieved 67.2% (± SD). Precision, recall, and F1-scores indicate moderate but stable predictive performance across folds. These results are consistent with heterogeneous multi-country crash datasets and support the study's emphasis on interpretable risk pattern identification rather than operational deployment.

3.3 Feature Importance and SHAP Interpretation

Random Forest rankings revealed distinct severity mechanisms across user groups. For micromobility users, age was the most influential predictor, followed by weather, lighting, and spatial context. SHAP analysis confirmed increased predicted severity among younger and older riders. Weather and reduced visibility also contributed to higher risk, while observations labelled as "unknown" were excluded from interpretation.For other road users, the rural context was the dominant determinant, followed by male gender and adverse weather. These findings indicate structurally different vulnerability profiles, with micromobility severity more strongly linked to user characteristics and other users more affected by environmental context.

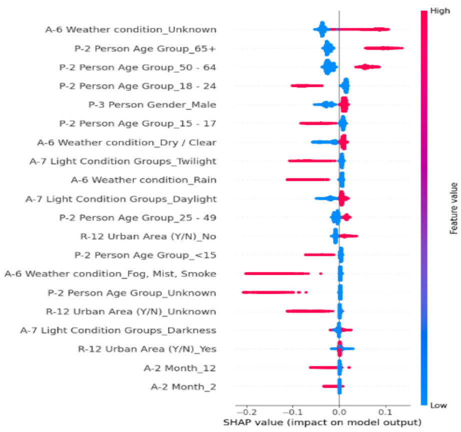

Fig. 2. SHAP Summary Plot for micromobility

3.4 Discussion and Policy Implications

This study highlights three core insights. First, micromobility riders' risks are shaped primarily by age and personal vulnerability, with both younger and older riders more prone to severe injuries. Second, environmental and contextual factors differ across groups, with the rural context being more decisive for other road users. Third, interpretable machine learning provides policy-relevant insight beyond predictive accuracy alone.

These findings suggest several targeted interventions. For micromobility users, measures should include youth-focused education, infrastructure adapted for older riders, and visibility or weather-related safety enhancements. For other road users, rural-focused strategies are critical. In both groups, addressing male risk-taking behavior remains an important dimension. By combining descriptive trends with interpretable machine learning, the analysis provides actionable, mode-specific insights to guide more effective and equitable road safety strategies.

4 Conclusion

This study integrated longitudinal trend analysis and interpretable machine learning to examine injury severity in micromobility-related crashes across Europe. Using standardized 2022 crash data, the results revealed distinct severity mechanisms between micromobility and other road users. Age emerged as the dominant predictor for micromobility injury severity, highlighting the importance of vulnerability-oriented safety policies. In contrast, contextual and environmental factors, particularly rural location and weather conditions, were more influential for other road users. The models demonstrated moderate but stable predictive performance, supporting the value of interpretable approaches in complex and heterogeneous datasets.

Overall, the proposed framework contributes to the growing body of research on explainable artificial intelligence in road safety and supports the development of targeted, evidence-based interventions for different user groups.

4.1 Limitations and Future Work

Several limitations should be considered. Despite harmonization, cross-country differences in reporting and injury coding may affect generalizability. The binary severity classification enhances interpretability but reduces information, and future studies should explore multi-class modelling approaches. The analysis was limited to structured variables, while unstructured data such as spatial imagery or crash narratives could improve contextual understanding. Although SHAP enhances transparency, it does not fully capture interactions or temporal dynamics. Future research should extend the framework to additional years, incorporate temporal modelling, and evaluate real-time risk prediction and policy simulation. Finally, although 2022 was contextualized within longitudinal trends, single-year predictive modelling may still reflect short-term fluctuations.

5 References

1. Hosseinzadeh, A., Karimpour, A., Kluger, R.: Factors influencing shared micromobility services: An analysis of e-scooters and bikeshare. Transportation Research Part D: Transport and Environment 100 (2021). https://doi.org/10.1016/j.trd.2021.103047
2. Ferenchak, N.N., Marshall, W.E.: Traffic safety for all road users: A paired comparison study of small & mid-sized U.S. cities with high/low bicycling rates. Journal of Cycling and Micromobility Research 2, 100010 (2024). https://doi.org/10.1016/j.jcmr.2024.100010
3. Prencipe, L.P., De Bartolomeo, S., Caggiani, L., Ottomanelli, M.: An injury severity- based methodology for assessing priority areas for shared micromobility accident risk mitigation. Travel Behaviour and Society 39 (2025). https://doi.org/10.1016/j.tbs.2024.100962
4. Qawasmeh, B., Oh, J.S., Kwigizile, V.: Micro-mobility safety assessment: Analyzing factors influencing the micro-mobility injuries in Michigan by mining crash reports. Future Transportation 4(4), 1580–1601 (2024). https://doi.org/10.3390/futuretransp4040076
5. İnaç, H.: Micro-mobility sharing system accident case analysis by statistical machine learning algorithms. Sustainability 15(3) (2023). https://doi.org/10.3390/su15032097
6. Sanjurjo-de-No, A., Pérez-Zuriaga, A.M., García, A.: Analysis and prediction of injury severity in single micromobility crashes with random forest. Heliyon 9(12) (2023). https://doi.org/10.1016/j.heliyon.2023.e23062
7. Abdi, A., O'Hern, S.: Understanding e-scooter rider crash severity using a built environment typology: A two-stage clustering and random parameter model analysis. Accident Analysis and Prevention 215 (2025). https://doi.org/10.1016/j.aap.2025.108018
8. Sohaee, N., Azadjoo Tabari, A., Sardari, R.: A multilevel spatial framework for e-scooter collision risk assessment in urban Texas. Safety 11(3), 67 (2025). https://doi.org/10.3390/safety11030067
9. Sarker, M.A.A., Asgari, H., Chowdhury, A.Z., Jin, X.: Exploring micromobility choice behavior across different mode users using machine learning methods. Multimodal Transportation 3(4) (2024). https://doi.org/10.1016/j.multra.2024.100167
10. Sadeghi, M., Aghabayk, K., Quddus, M.: A hybrid machine learning and statistical modeling approach for analyzing the crash severity of mobility scooter users considering temporal instability. Accident Analysis and Prevention 206 (2024). https://doi.org/10.1016/j.aap.2024.107696