Available online at www.sciencedirect.com

ScienceDirect

Transportations Research Procedia 00 (2026) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

Road Safety and Simulation 2026 – RSS2026

Interpretable Machine Learning for Municipal Road Safety: A Spatiotemporal Analysis of Crash Severity in Athens (2016–2020)

Eleni Maria Theodoraki[a]*, Paraskevi Koliou[a], George Yannis[a]

[a]*National Technical University of Athens, Iroon Polytechniou 5-9, 15772 Athens, Greece*

Abstract

Urban accidents and crashes are one of the major causes of death and injuries in metropolitan areas like Athens, Greece, where intense traffic strain and urban dense infrastructure increase exposure risk. In this research, we develop an interpretable ensemble **machine learning approach** to understand and interpret the severity of crashes, incorporating data from 2016 to 2020 provided by **Smart Maps** and city traffic records. **Random Forests** and **XGBoost** models were developed based on an integrated dataset that includes **spatiotemporal, geo-infrastructural, and crash-related features**. The evaluation of the models showed strong predictive capabilities, with XGBoost slightly outperforming Random Forests on all predictive accuracy indicators.

To improve the model's interpretability and transparency, we applied **SHAP (SHapley Additive exPlanations)** analysis to identify the most important crash severity predictor variables. After applying SHAP analysis to the best-performing model, it was discovered that the most significant metrics were crash severity, mean fatalities, and longitude (caplon). Further, analysis through **Partial Dependency** and SHAP revealed the nonlinear and **spatial interactive effects** of urban configuration/ traffic and spatial crashes.

Subsequent use of Smart Maps geolocation data for spatial risk mapping depicted high-risk municipalities concentrated in central and southern Athens. Such spatial knowledge offers a decision-advocacy tool for urban planners and policymakers, helping them to focus on road safety improvements and optimize road infrastructure in targeted municipalities.

This study shows the potential of interpretable ensemble learning, coupled with Smart Maps geospatial intelligence, to develop advanced analyses of road safety for municipalities and to take proactive management of crash risk.

Keywords: *Road Safety; Random Forest; XGBoost; Machine Learning; SHAP; Spatial Analysis; Descriptive Analysis; Spatiotemporal modeling*

---

* Corresponding author. Tel.: + 6980438539
E-mail address: e_theodoraki@msil.ntua.gr

1. Introduction

A critical challenge to public safety globally is the existence of road traffic crashes, which lead to a significant rise in fatalities and injuries, causing approximately 1.35 million deaths and 50 million injuries each year. (Chen et al., 2025;Xiao & Duan, 2025;Dong et al., 2022). Road fatalities have reached and almost outpaced cancer death rates as the sixth global cause of death by the next decade (Aziz et al., 2024). However, there has been a noticeable reduction in the annual number of road traffic deaths, down to 1.19 million. Transportation agencies worldwide, including the Federal Highway Administration (FHWA), have set a goal to eliminate road traffic fatalities (Islam et al., 2024). Despite decades of research, the severity of crashes in urban areas, such as Athens, is a difficult aspect to predict due to the combination of spatiotemporal, roadway, demographic, and other variables.

Traditional statistical models provide interpretability but often struggle to capture nonlinear and high-dimensional crash dynamics. Machine learning approaches improve predictive performance; however, their limited transparency restricts their practical use in municipal planning. At the same time, crash severity exhibits strong spatial heterogeneity, as local land use patterns, infrastructure, and traffic demand vary across municipalities.

This study proposes an interpretable ensemble machine learning approach to predict the severity of municipal crashes in Athens (2016-2020) using random forest and XGBoost, SHAP-based interpretable machine learning, and geospatial risk mapping using Smart Maps data. This research uses this framework to provide interpretable predictions that further spatially target urban road safety planning. We conclude that XGBoost has slightly improved prediction power over Random Forest. SHAP analysis suggests that spatial position (and specifically, longitude) and, crash intensity indicators are the most powerful predictors of crash severity, and risk maps at the municipality level can be a useful tool for proactive measures.

2. Literature Review

2.1 The Necessity of Urban Crash Severity Analysis

Accidents that occur in urban metropolitan areas are a major concern, as they disrupt the sustainable operation of traffic flow and cause severe congestion (Parsa et al., 2020; Shi et al., 2025). A central objective in road safety is the accurate prediction and analysis of the severity of road accidents, which is essential for the definition of preventive measures and enhancing road safety (Hamdan & Sipos, 2025;Aziz et al., 2024). Conventional statistical models, such as those based on the logit or probit function, often face challenges in modeling the complex, nonlinear relationships that feature in real-world crash statistics (Hamdan & Sipos, 2025;Dong et al., 2022; Chen et al., 2025). Consequently, there was a marked transition towards more flexible, data-driven models, such as machine learning algorithms, which lend themselves well to handling complex, high-dimensional crash statistics (Kashifi et al., 2023)

2.2 Interpretable Ensemble Machine Learning

Ensemble methods are widely valued for their superior predictive capabilities compared to single statistical or machine learning approaches (Hamdan & Sipos, 2025). **XGBoost** is considered an efficient implementation of gradient-boosted decision trees. It generally provides high accuracy and processing speed, while being computationally inexpensive compared to other techniques (Parsa et al., 2020). For example, in studies, it's observed that XGBoost achieved high accuracy in the prediction, such as 86,73% while analyzing the older pedestrian crashes, and 89.05% when combined with a Bayesian network for freeway crash severity (Hamdan & Sipos, 2025). **Random Forests** is also a very efficient ensemble learning technique that constructs multiple decision trees at training time. This algorithm repeatedly arises as a top-performing algorithm in the prediction of accident severity across different datasets due to its resistance to overfitting and noise (Çeven & Albayrak, 2024;Yan & Shen, 2022). In the study of (Hamdan & Sipos, 2025) on 3.5 million crash records in the U.S., Random Forests was confirmed to be the best model, achieving 97,4% accuracy.

2.3 Interpretable Machine Learning and SHAP Analysis

A major challenge in adopting Machine Learning models for transportation safety is their lack of transparency. This challenge arises because machine learning (ML) models are often considered "black box" models, making it difficult to interpret their results (Parsa et al., 2020). Recently, there has been a growing convergence of studies that address explainable-AI (XAI) models in predicting crash severities (Li, 2022). SHapley Additive exPlanations, based on cooperative game theory, provides reliable and locally sound attributions of feature contributions with explainable boosted-tree models (Aziz et al., 2024;Xiao & Duan, 2025). SHAP was applied in elucidating the complex nonlinear relationships of variables such as road geometry, weather, demographics, and temporal variables in predicting crash severities, many of which were not considered through traditional models (Parsa et al., 2020). Other explanatory models, such as Partial Dependence Plots (PDP) tools, explain the model's global properties, elucidating monotonic or nonlinear relationships (Li, 2022).

2.4 Spatial Heterogeneity and Geospatial Intelligence in Crash Modelling

Crash risk is considered spatial in nature, with built environment patterns, corridor shapes, land use, and population activity

Peer-review under responsibility of the scientific committee of the Euro Working Group on Transportation Annual Meeting 2025 - EWGT2025

being some of its determinants (Guo et al., 2024). Studies carried out using coordinate representations or Geographical Information System (GIS) variables indicate the importance of spatial elements as some of the most dominant variables for modeling crash outcome predictions(Li, 2022). Recent applications of Machine Learning models indicate the use of real geospatial information (Kashifi et al., 2023), such as urban GIS, map variables, or telemetry information, along with powerful, explainable models, like Interpretive Models, for risk zone mapping for roads, intersection points, and even entire neighborhoods (Xiao & Duan, 2025).

While ensemble learning and explainable AI have been widely studied, fewer works integrate interpretable machine learning with fine-grained municipal geospatial data for urban crash severity analysis. This study addresses this gap by combining Random Forest, XGBoost, SHAP explanations, and municipality-level spatial risk mapping in Athens.

3. Methodology

3.1 Data Representation

The dataset contains crash records in different municipalities of Athens for the period 2016-2020 and was extracted using the **SmartMaps** geospatial infrastructure. Every row is a representation of a reported crash and is characterized by the following set of features:

- **Crash details**, such as severity level, crash type, number of vehicles, and vulnerable road users involved.
- **Temporal attributes**: Year, Month, Day of Week, and Hour of the Day.
- **Environmental attributes**: lighting, weather, road conditions.
- **Spatial/geographic variables**: latitude, longitude, municipality identifier, and GIS variables (road density, closeness to intersections, land use).

All variables were incorporated into a tabular format optimal for machine learning. Categorical variables were encoded using target encoding or one-hot encoding, depending on the number of categories, and numerical variables were scaled using standardization or min-max scaling. Spatial variables were left unaltered for SHAP analysis and mapping.

The study also carried out a missing data audit. Incomplete severity result records, which were less than 2% of the total, were excluded, and missing environmental variables were treated by using medians at the municipality level. The constructed dataset applies to predictive modelling and spatial interpretation. Municipal-level descriptive grouping was explored using hierarchical clustering for exploratory purposes; results were used only to support spatial interpretation. The final dataset comprised 119 crash records after excluding incomplete severity entries (< 2% of total).

The severity variable was distributed as follows:

- Slight injury: 94.6% of records
- Serious injury: 2.8% of records
- Fatal: 2.6% of records

This distribution was used to stratify the 80/20 train-test split and to apply class weights in the Random Forest model.

3.2 Statistical Modelling

To explore the changes over time and structure regarding crash severity patterns, two types of statistical analysis were conducted:

The first type of statistical analysis was the **T-test analysis.** To assess mean severity scores between pre- and post-2020, a two-sample t-test analysis by Welch's t-test is conducted. This assessment evaluates whether macro-level changes, such as changes to mobility and pandemic-related behavioral changes impacting travel, influence statistical differences.

The second analysis conducted was a **Linear Regression** to identify long-term trends in specific categories of crash severity:

$\text{Severity}_t = \beta 0 + \beta 1 \cdot \text{year} + \epsilon$

In this case, the slope $\beta 1$ represents direction and rate of change regarding severity over time, and R^2 represents the predictability of severity over time. These statistics were then compared to machine learning-based feature attribution maps to ensure consistency with machine learning analysis.

3.3 Machine Learning

In this section, a basic predictive framework is given that incorporates techniques of ensemble learning, SHAP interpretability, and geospatial risk mapping. Firstly, the data was divided into 80% and 20% subsets, which served as training and testing sets, and it was stratified based on severity level to ensure class balance. For robustness, other analysis methods used 5-fold cross-validation.

For the model training, two models were developed:

1. **Random Forest (RF)**: 500 trees, equal class weights, bootstrap.
2. **XGBoost**: Gradient Boosted Decision Trees, with optimal hyperparameters (learning rate, maximum tree depth, subsampling)

Peer-review under responsibility of the scientific committee of the Euro Working Group on Transportation Annual Meeting 2025 - EWGT2025

XGBoost used early stopping to prevent overfitting. Hyperparameter tuning was conducted by **Bayesian** optimization. Model performance was evaluated using R^2 (explained variance), MAE (mean absolute error), RMSE (root mean square error), and CV-R^2 (cross-validated R^2 from 5-fold cross-validation) to assess both in-sample fit and out-of-sample generalisability. These metrics together quantified classification and fairness with respect to spatial units.

To maintain transparency and insight, the top-performing model was rendered interpretable using TreeSHAP, a method that calculates exact Shapley-value explanations for ensembles of decision trees.

Global Interpretability

- SHAP summary plots show which predictors dominate.
- SHAP dependence plots uncover nonlinear effects.
- SHAP interaction values expose how spatial factors and crash-related variables relate to each other.

Local Interpretability

- Municipality-level SHAP profiles are informative about which features raise or lower the severity of certain areas.
- Crash-specific explanations allow for case-by-case understanding.

It is this layer of interpretability that turns the ML model into a practical tool for policy and urban safety planning.

To translate the model outputs into usable geospatial intelligence, a spatial risk index was calculated for each municipality by aggregating three components:

1. the mean predicted severity score from the XGBoost model across all crash records within the municipality
2. the mean SHAP contribution of spatial features (longitude, latitude, and GIS density variables) for that municipality, reflecting how much spatial position drives predicted severity
3. the empirical severity rate, calculated as the ratio of KSI (killed or seriously injured) crashes to total crashes in the municipality over 2016-2020.

The three components were normalised to a 0-1 scale and combined using equal weights into a composite index, which was then mapped using SmartMaps geospatial layers to identify high-risk municipalities.

4. Results

This section presents the results of the descriptive, spatial, and machine learning analysis applied to the crash dataset for different municipalities of Athens during the period 2016-2020.

4.1 Descriptive analysis

Figure 1 illustrates the temporal analysis of road crashes, fatalities, serious injuries, and slight injuries. It's observed that there is a stable pattern in crashes and fatalities from 2016 to 2019 and a sharp drop in 2020. This drop is due to the mobility restrictions of COVID-19, resulting in a reduction of risk exposure. Slight injuries contributed most to the casualties. Deaths were relatively stable over the five years.

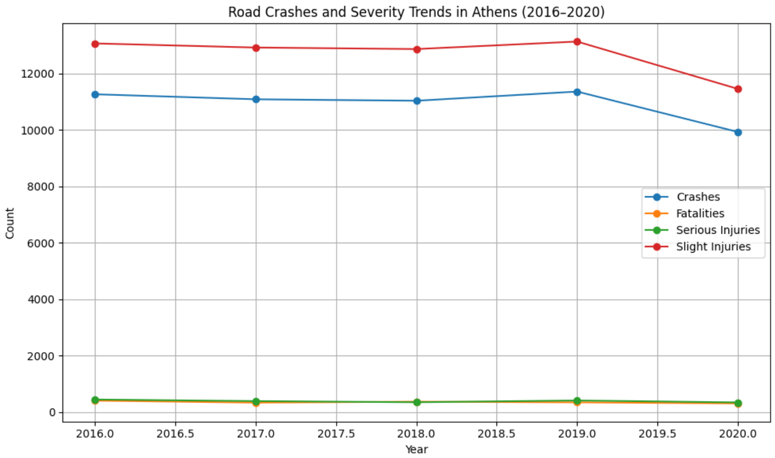
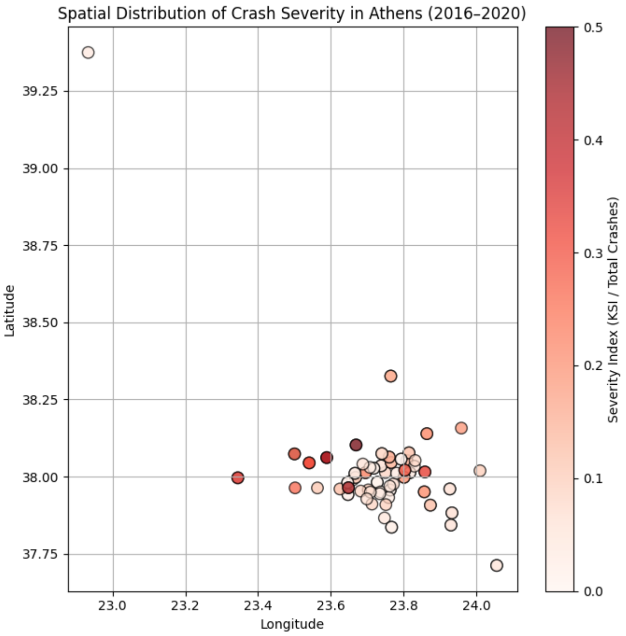


Figure 1: *Temporal Trends in Crashes and Fatalities*

Region-wise, the number of crashes took place in central municipalities with a large amount of traffic and a high level of activity, as presented in **Figure 2**. In **Figure 3,** the comparison of road types is presented, and it is found that a larger number of crashes and injuries were identified in non-motorway sections.

Peer-review under responsibility of the scientific committee of the Euro Working Group on Transportation Annual Meeting 2025 - EWGT2025



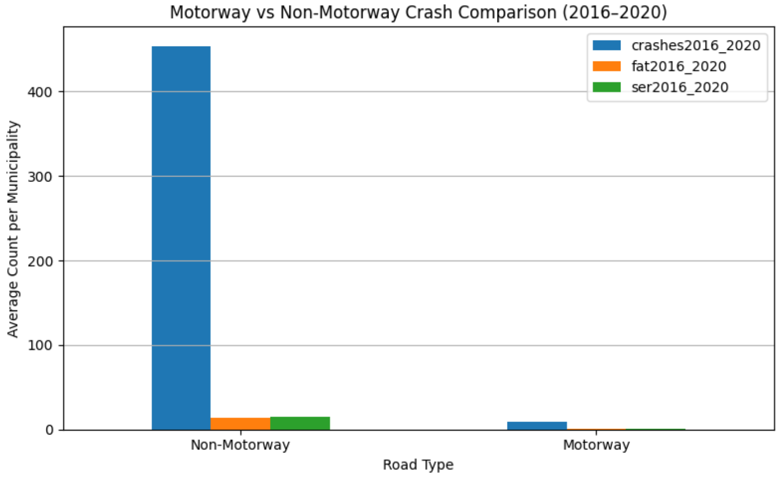


Figure 2: *Crash Density Map of Athens (2016–2020)*

Figure 3: *Motorway vs Non-Motorway Crash Comparison*

4.2 Spatial Patterns

The mapping of the severity index in Athens, which is presented in **Figure 5**, revealed a strong spatial variability:

- **Western municipalities**, such as Aspropyrgos, Elesfina, Megara, and Salamina, had a great level of severity with relatively low numbers.
- **Central municipalities** had a high rate of crashes but a lower ratio of severity.

Inversely correlated data emphasize a need for models with spatial awareness for predictions and explain why geo-spatial variables later became major predicting factors.

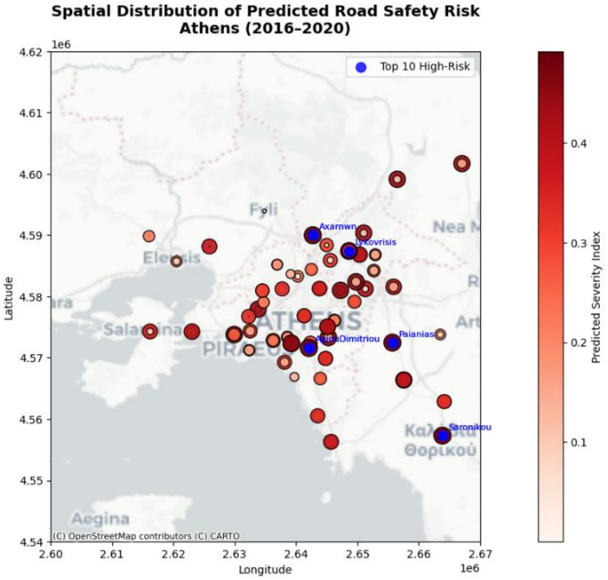


Figure 4: *Spatial Distribution of Historical Crash Severity*

4.3 Model Performance

Two machine learning models were developed, Random Forests and XGBoost, and were trained to predict the crash severity of every municipality. **Table 1** summarizes the performance of the models.

Table 1. Model Performance Metrics

Model	**R^2**	**MAE**	**RMSE**	**CV-R^2**
Random Forest	0.296	0.06197	0.11008	0.4238
XGBoost	**0.302**	0.06261	**0.10961**	**0.5499**

It is observed that **XGBoost** achieved the **highest performance**, R^2=0.302 and CV-R^2=0.5499, outperforming Random Forest in all evaluation metrics but MAE. While R^2 values of approximately 0.30 are consistent with published benchmarks for aggregate crash-severity models at the municipal or zone level, where unexplained variance is expected due to unobserved factors such as driver behaviour, real-time traffic volume, and enforcement (Kashifi et al., 2023; Li, 2022; Guo et al., 2024). The cross-validated R2 of 0.55 for XGBoost indicates that the model generalises well beyond the training data, which is the primary objective for a spatial risk-mapping tool. The graph of predicted vs. actual values for severity (**Figure 6**) shows a good fit and verifies that the model captures the key factors for municipal levels of severity.

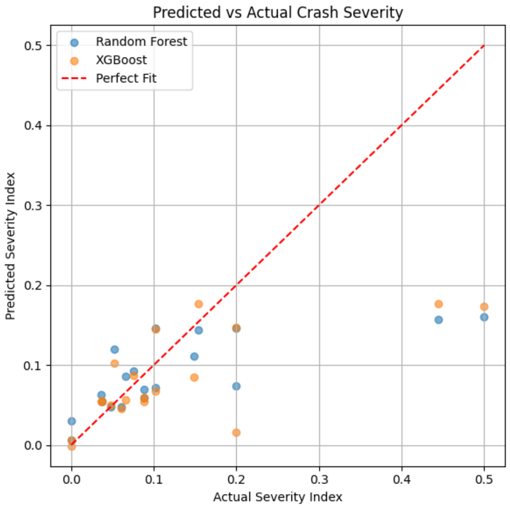


Figure 5: Predicted vs Actual Crash Severity

4.4 SHAP Analysis and Feature Importance

The SHAP analysis was developed for the XGBoost model and quantifies the importance of features. In **Table 2**, we summarize the top contributors.

Table 2. SHAP Feature Importance

Feature	**Mean**
caplon (longitude)	**0.0353**
mean_fatalities	0.0246
mean_crashes	0.0192
mean_slight	0.0160
motorway	0.0146
ksi_total	0.0104
mean_serious	0.0049
lon_zscore	0.0042
crash_change_rate	0.0037

Peer-review under responsibility of the scientific committee of the Euro Working Group on Transportation Annual Meeting 2025 - EWGT2025

Feature	**Mean**
fatality_change_rate	0.0033
caplat (latitude)	0.0022
lat_zscore	0.00037

From the SHAP analysis, we can draw some key findings. Firstly, the **Longitude (caplon)** is the most important feature, confirming that **east-west spatial variation** is the strongest predictor for severities. Furthermore, features such as mean fatalities, mean crashes, and mean slight injuries capture exposure and casualty intensity. Municipalities with a larger proportion of motorway segments have shown types of severity profiles. Lastly, the role of derived z-score spatial features was secondary (lon_zscore and lat_zscore).

The SHAP dependencies plots, presented in **Figures 6,7**, show nonlinear dependencies, especially for:

- **Caplon:** The level of severity certainly escalates in Western municipalities.
- **Mean fatalities:** Nonlinear escalation for municipalities with an average of more than 1 fatality.
- **Mean crashes:** Diminishing returns as the crash volume increases.

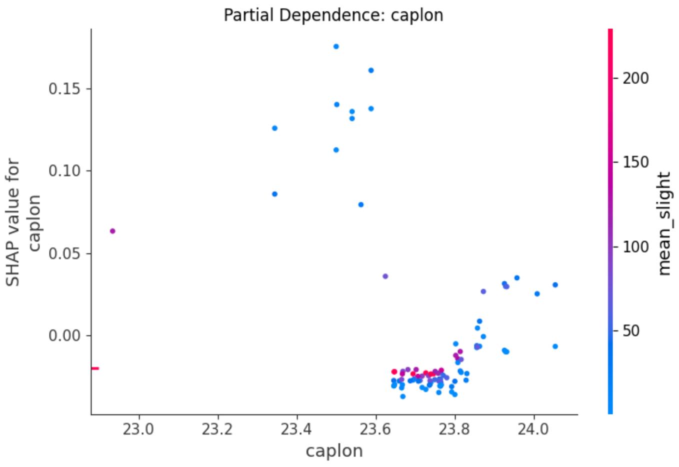


Figure 6: Partial dependence: caplon

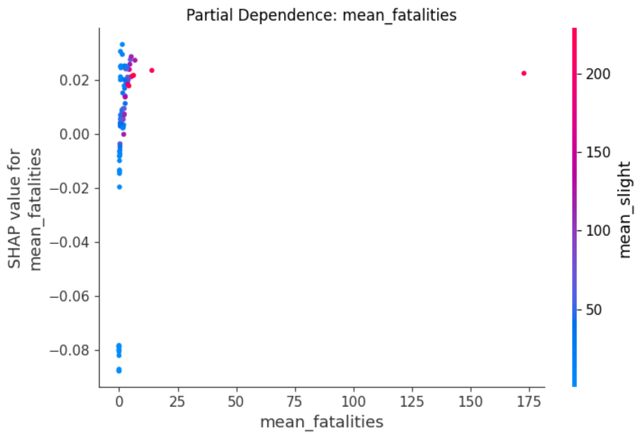


Figure 7: Partial dependence: mean fatalities

5. Discussion

The results of the research point out the strong spatial and contextual dependence of the severity of crashes across the municipalities of Athens. The western municipalities, such as Aspropyrgos, Elesfina, and Salamina, systematically exhibit higher levels of severity irrespective of lower frequencies. This tends to indicate that particular conditions exist in those places. The contrast between municipalities with many crashes but low levels of severity and municipalities with low numbers of crashes but large levels of severity illustrates the importance of a priority focused on severity instead of volume. The explainability of SHAP values was critical because complex phenomena such as nonlinear thresholds for fatality, interactions between locations, and contextual variables such as motorway presence were explained. Despite a modest level of predictability, well within expectations for a municipal aggregate model, the explanations and maps represent a great asset concerning municipal planning for road safety. 'It should be noted that the spatial heterogeneity patterns identified in this study may be sensitive to the inclusion of additional variables such as traffic volume, land use, and population density, which were not available for this analysis. Future work incorporating these variables may reveal different or more nuanced spatial patterns. Thus, foremost among its uses would be as a tool for defining which municipalities would be in need of intervention and for what reason.

6. Conclusion

This research offers a machine learning model with explanations for predicting the level of crashes in urban areas for the city of Athens using methods such as ensemble learning and SHAP explanations. The model uses traditional data for crashes along with spatial variables from SmartMaps to provide predictions with explanations.

Peer-review under responsibility of the scientific committee of the Euro Working Group on Transportation Annual Meeting 2025 - EWGT2025

XGBoost has the best overall performance, and SHAP values highlighted that spatial location, especially longitude, is the most influential factor in evaluating municipal road crash severity. The data clearly show distinct differences in road crash severity across Western municipalities that appear more severe than would be expected based on their number of crashes. This clearly indicates that a more sophisticated approach to managing road safety needs to be employed.

The components for interpretability were critical for extracting nonlinear relationships that would have remained obscure using conventional approaches. The spatial risk map produced using this model serves as a useful tool for hotspot detection and assisting municipalities in dealing with hotspots.

The research recognizes the following limitations:

- **Data Granularity**: The prediction of severity was done at a municipal level; the data may be more accurate for a road segment.
- **A lack of diversity in feature space**: Variables such as traffic volume data, socioeconomic factors, or enforcement data would be very beneficial, but were unavailable.
- **Temporal evolution**: The model does not currently consider changes in mobility in urban environments (e.g., COVID-19 pandemics, modal shift changes over time).

Future studies may consider:

- Fusion of multimodal data sources (traffic volume data, land use composition data, public transportation data).
- Creating spatiotemporal models for disease severity with annual changes.
- Development of a decision-making interface for real-time urban risk observation based on this framework.
- Studying deep learning models or graph-based spatial models while retaining interpretability.

7. References

Aziz, K., Chen, F., Khan, I., Hussain Khahro, S., Malik Muhammad, A., Ahmed Memon, Z., & Khattak, A. (2024). Road Traffic Crash Severity Analysis: A Bayesian-Optimized Dynamic Ensemble Selection Guided by Instance Hardness and Region of Competence Strategy. *IEEE Access*, *12*, 139540–139559. https://doi.org/10.1109/ACCESS.2024.3465489

Çeven, S., & Albayrak, A. (2024). Traffic accident severity prediction with ensemble learning methods. *Computers and Electrical Engineering*, *114*. https://doi.org/10.1016/j.compeleceng.2024.109101

Chen, F., Liu, X. Q., Yang, J. J., Liu, X. K., Ma, J. H., Chen, J., & Xiao, H. Y. (2025). Traffic accident severity prediction based on an enhanced MSCPO-XGBoost hybrid model. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-00797-7

Dong, S., Khattak, A., Ullah, I., Zhou, J., & Hussain, A. (2022). Predicting and Analyzing Road Traffic Injury Severity Using Boosting-Based Ensemble Learning Models with SHAPley Additive exPlanations. *International Journal of Environmental Research and Public Health*, *19*(5). https://doi.org/10.3390/ijerph19052925

Guo, M., Janson, B., & Peng, Y. (2024). A spatiotemporal deep learning approach for pedestrian crash risk prediction based on POI trip characteristics and pedestrian exposure intensity. *Accident Analysis and Prevention*, *198*. https://doi.org/10.1016/j.aap.2024.107493

Hamdan, N., & Sipos, T. (2025). Advancements in Machine Learning for Traffic Accident Severity Prediction: A Comprehensive Review. In *Periodica Polytechnica Transportation Engineering* (Vol. 53, Issue 3, pp. 347–355). Budapest University of Technology and Economics. https://doi.org/10.3311/PPtr.40369

Islam, R., Abdel-Aty, M., Wang, D., & Islam, Z. (2024). Spatial Ensemble Distillation Learning for Large-Scale Real-Time Crash Prediction. *IEEE Transactions on Intelligent Transportation Systems*, *25*(11), 16506–16521. https://doi.org/10.1109/TITS.2024.3432854

Kashifi, M. T., Al-Turki, M., & Sharify, A. W. (2023). Deep hybrid learning framework for spatiotemporal crash prediction using big traffic data. *International Journal of Transportation Science and Technology*, *12*(3), 793–808. https://doi.org/10.1016/j.ijtst.2022.07.003

Li, Z. (2022). Extracting spatial effects from machine learning model using local interpretation method: An example of SHAP and XGBoost. *Computers, Environment and Urban Systems*, *96*. https://doi.org/10.1016/j.compenvurbsys.2022.101845

Parsa, A. B., Movahedi, A., Taghipour, H., Derrible, S., & Mohammadian, A. (Kouros). (2020). Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. *Accident Analysis and Prevention*, *136*. https://doi.org/10.1016/j.aap.2019.105405

Xiao, Y., & Duan, Z. (2025). An explainable multi-task deep learning framework for crash severity prediction using multi-source data. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-09226-1

Yan, M., & Shen, Y. (2022). Traffic Accident Severity Prediction Based on Random Forest. *Sustainability (Switzerland)*, *14*(3). https://doi.org/10.3390/su14031729

Peer-review under responsibility of the scientific committee of the Euro Working Group on Transportation Annual Meeting 2025 - EWGT2025