Periodica Polytechnica Transportation Engineering, 53(3), pp. 347–355, 2025

Advancements in Machine Learning for Traffic Accident Severity Prediction: A Comprehensive Review

Noura Hamdan[1], Tibor Sipos[1,2*]

[1] Department of Transport Technology and Economics, Faculty of Transportation Engineering and Vehicle Engineering, Budapest University of Technology and Economics, Műegyetem rkp. 3., H-1111 Budapest, Hungary
[2] KTI - Hungarian Institute for Transport Sciences and Logistics Ltd., Than Károly str. 3–5., H-1119 Budapest, Hungary
[*] Corresponding author, e-mail: sipos.tibor@kjk.bme.hu

Received: 06 March 2025, Accepted: 30 May 2025, Published online: 13 June 2025

Abstract

This literature review categorizes machine learning studies in traffic accident severity prediction, providing a comprehensive overview of the diverse applications and advancements in this field. It begins with a comparative analysis of machine learning models, highlighting the performance of various algorithms such as Random Forest, XGBoost, and Support Vector Machines (SVM) in predicting accident severity. The review also explores factor-specific studies, emphasizing the influence of road, environmental, and vehicle-related factors on crash outcomes. These studies demonstrate the critical role of factors such as road type, weather conditions, and vehicle characteristics in determining accident severity. Additionally, crash-type-specific prediction models have been developed, showcasing the ability of machine learning models to tailor predictions based on the nature of the crash, whether involving pedestrians, vehicles, or specific collision types. The review also examines hybrid and ensemble approaches, which combine multiple algorithms to enhance prediction accuracy. These approaches leverage the strengths of individual models to improve overall performance, offering a promising direction for future research. By categorizing the studies into these key areas, this review provides a structured understanding of the state-of-the-art in machine learning applications for traffic accident severity prediction and identifies opportunities for further development to enhance prediction robustness, accuracy, and applicability in real-world traffic safety management.

Keywords

traffic accident severity, machine learning models, comparative analysis, hybrid approaches, prediction models

1 Introduction

As the global momentum toward sustainable development intensifies, the intersection of traffic safety and sustainability has emerged as an area of critical importance (Török, 2017). The application of machine learning (ML) models to predict traffic accident severity offers substantial potential to mitigate fatalities and injuries, while simultaneously contributing to the development of safer, more resilient transportation systems that promote long-term environmental and social sustainability (Hee et al., 2024). A comprehensive understanding and accurate prediction of traffic accident severity is essential for implementing targeted preventive measures and reducing fatalities on the road (Abdullah and Sipos, 2022). The severity of injuries resulting from road accidents is influenced by a multifaceted interaction of factors, including driver behavior, road infrastructure, vehicle characteristics, and environmental conditions (Jima and Sipos, 2022; Ötvös and Török, 2024). Identifying and comprehending these factors is pivotal to enhancing road safety and enacting effective interventions aimed at mitigating crash severity (Sipos et al., 2021; Wang et al., 2023; Zou et al., 2021).

Historically, traffic injury severity analysis has predominantly relied on statistical methods such as logistic regression and ordered probit models (Garrido et al., 2014; Rifaat and Chin, 2007). While these models have provided valuable insights, they often struggle to capture the complex, non-linear relationships inherent in real-world crash data. However, with advancements in computational techniques and the expanding availability of large-scale traffic datasets, machine learning algorithms have surfaced as a promising alternative (Gururaj et al., 2022). These models offer superior predictive capabilities, adept at uncovering intricate patterns and interactions between variables, thereby enhancing the accuracy of injury severity predictions.

Cite this article as: Hamdan, N., Sipos, T. (2025) "Advancements in Machine Learning for Traffic Accident Severity Prediction: A Comprehensive Review", Periodica Polytechnica Transportation Engineering, 53(3), pp. 347–355. https://doi.org/10.3311/PPtr.40369

This study presents a comprehensive review of the application of machine learning in modeling road crash injury severity. The aim is to provide researchers and policymakers with a nuanced understanding of the effectiveness of various ML approaches, while identifying key challenges and suggesting future avenues for research. This paper presents state of the art studies that compare different analytical methods applied in injury severity prediction in road traffic crashes. To do so, three research questions were defined:

- Did the study develop models for traffic crash injury severity prediction?
- How were the performances of the different models compared?
- Which model achieved the best performance?

The search engines implemented were Scopus (Elsevier, online), ScienceDirect (Elsevier, online), and IEEE Xplore (IEEE, online). Additional articles were identified analyzing the citations referenced in other articles. The literature review was performed from 2020 until February of 2025 with a total of 32 studies. Search terms included crash, accident, injury severity, prediction models, and machine learning. These search terms were based on what is commonly found in the literature.

2 The applications of the Machine Learning algorithms for road crash injury severity prediction

For each of the studies reviewed, the following elements are discussed: the analytical methods employed, the performance metrics used for comparison, and the algorithms that were reported to yield the best results. The studies reviewed in this paper utilize datasets from various countries, including Europe, India, China, Canada, Saudi Arabia, South Africa, Colombia, Türkiye, Taiwan, and the United States, with the latter contributing the highest number of studies (12). However, the quality and structure of these datasets vary significantly, ranging from aggregated crash reports to real-time sensor and GPS data, which may impact model performance and generalizability. Despite the global distribution of these studies, cross-regional comparisons remain challenging due to differences in accident reporting standards, traffic laws, and road infrastructure. Consequently, the transferability of machine learning models across different regions is not always straightforward. Fig. 1 presents a summarized methodology commonly adopted in the reviewed studies, outlining key steps such as data collection, preprocessing, feature selection, model training and evaluation, and interpretability techniques.

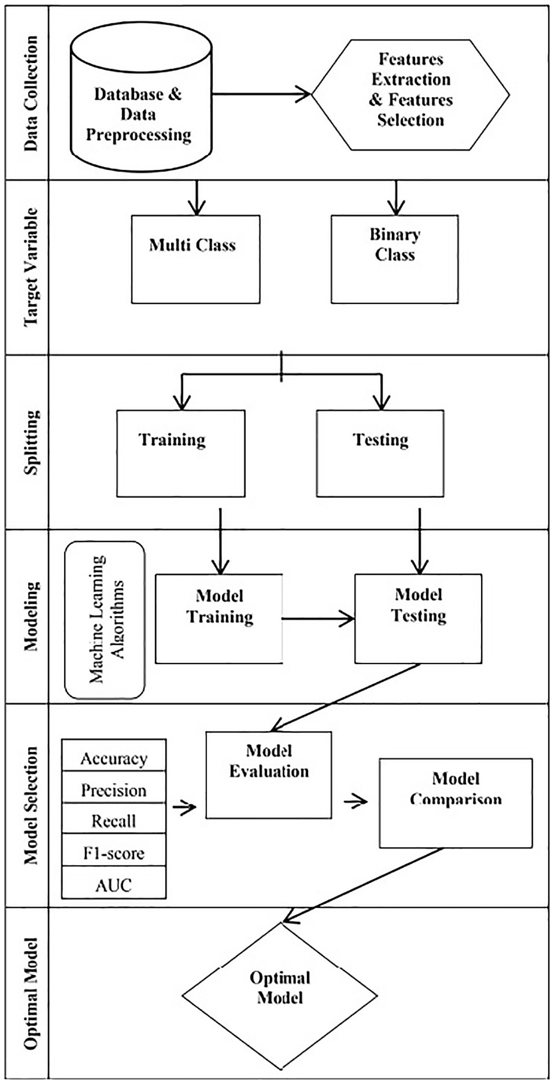

Fig. 1 Overview of the methodological framework used in the reviewed studies

Additionally, the review highlights the most frequently used and most effective ML techniques across the studies, providing a brief overview of those that have shown the most promise in predicting injury severity. Table 1 shows the Number of times that the different prediction methods were used in the scope of this literature review.

3 Categorization of Machine Learning studies in traffic accident severity prediction

3.1 Comparative analyses of Machine Learning models

The comparative evaluation of machine learning models in predicting traffic accident severity has become an essential area of research, shedding light on the relative effectiveness of various algorithms under diverse conditions. Such analyses are pivotal for understanding the nuances of model efficiency, accuracy, and interpretability, providing

Table 1 Number of times that the different prediction methods were used

Algorithm	Number of times used
Gradient boost	3
AdaBoost	6
LightGBM	4
Logistic Regression	8
K-Nearest Neighbor	6
Naïve Bayes	8
Random Forest	18
Artificial Neural Network	3
XGBoost	10
Multilayer Perceptron (MLP)	4
Decision Tree	6
Random Tree	2
Support Vector Machine (SVM)	8
Ordered Probit	2

crucial insights into how different models perform when applied to real-world accident datasets.

Several studies have demonstrated the superiority of Random Forest in predicting accident severity. Chen and Chen (2020) developed and compared Logistic Regression, Classification and Regression Tree (CART), and Random Forest models using data from Taiwan's highways. Their results indicated that Random Forest outperformed other models in accuracy, sensitivity, and specificity. Similarly, Umer et al. (2020) conducted an extensive study on 3.5 million crash records across 49 U.S. states, employing an ensemble of Logistic Regression, Stochastic Gradient Descent, Random Forest, AdaBoost, Extra Trees, and Gradient Boosting Machine. Their analysis confirmed Random Forest as the most effective model, achieving an accuracy of 97.4%. Likewise, Bokaba et al. (2022) analyzed machine learning classifiers, including Naïve Bayes, Logistic Regression, k-Nearest Neighbors, AdaBoost, Support Vector Machine, and Random Forest, using South African crash data. Their findings revealed that Random Forest, when combined with multiple imputations for handling missing data, exhibited the highest predictive accuracy.

Beyond these comparative studies, additional research has explored specific methodological advancements and contextual influences on model performance. These studies are summarized in Table 2.

This structured synthesis underscores the growing body of research exploring machine learning applications in traffic accident severity prediction. While Random Forest consistently emerges as a leading model, other advanced techniques, including ensemble methods and deep learning approaches, show promise in specific contexts. Furthermore, these studies emphasize the critical role of dataset characteristics, feature selection strategies, and domain-specific factors in determining model effectiveness.

Table 2 Summary of additional machine learning studies on traffic accident severity prediction

Study	Country of the dataset	Machine Learning models	Key findings
Infante et al. (2022)	Portugal	Statistical models, Machine Learning	ML models performed comparably to traditional methods in large datasets but struggled with small, imbalanced datasets.
Cicek et al. (2023)	USA	Decision Trees, Neural Networks (MLP), Support Vector Classifier, Case-Based Reasoning, Naïve Bayes	MLP achieved the highest accuracy; Shapley values identified seatbelt usage, alcohol consumption, and speed violations as critical factors.
Obasi and Benson (2023)	UK	Random Forest, Logistic Regression	Both models achieved 87% accuracy; highlighted challenges in generalizing machine learning results across regions.
Sorum and Pal (2024)	India	12 ML models, including LightGBM	LightGBM outperformed other models; accident cause, collision type, and vehicle type were the most influential factors.
Daoud et al. (2025)	USA	Logistic Regression, k-NN, Naïve Bayes, Random Forest, ANN, XGBoost, LSTM	Random Forest achieved 97.6% accuracy; study proposed a smartphone-based illumination data collection tool.

3.2 Applications of specific Machine Learning techniques

This category encompasses studies that focus on the application or enhancement of specific machine learning techniques for the prediction of accident severity. The selected studies demonstrate the diverse approaches and methodologies employed to improve the accuracy and reliability of traffic accident severity predictions across various datasets and contexts. Several studies have demonstrated the efficacy of eXtreme Gradient Boosting (XGBoost) in predicting accident severity. Shen and Wei (2020) analyzed hazardous material transport accidents in China, where XGBoost outperformed traditional statistical models in terms of classification performance. Their findings highlighted regional variations in accident severity determinants, emphasizing the necessity of localized safety measures. Similarly, Guo et al. (2021) employed XGBoost to analyze crash severity involving elderly pedestrians in Colorado, USA, achieving an accuracy of 86.73%. The study identified key risk factors, such as lighting conditions, pedestrian involvement, and rear-end collisions, stressing the need for targeted interventions.

Expanding on these findings, Yang et al. (2022) integrated XGBoost with a Bayesian network to model freeway crash severity, achieving a high prediction accuracy of 89.05%. Their study underscored the importance of feature interactions in safety management. Additionally, Zhao et al. (2022) utilized an interpretable XGBoost model to examine the impact of pavement friction on crash severity, applying SHAP (Shapley additive explanations) values to reveal nonlinear predictor interactions. Their findings emphasized that pavement friction below critical thresholds significantly increased accident severity, advocating for targeted road surface treatments.

Support Vector Machines (SVM) have also been widely explored for traffic accident severity prediction. Arhin and Gatiba (2020) evaluated eight SVM models with different kernel functions on crashes at unsignalized intersections in Washington, DC. Their results demonstrated that the radial basis function kernel achieved the highest accuracy (83.2%), whereas the Gaussian Naïve Bayes model performed the worst (48.5%), illustrating the significance of kernel selection in SVM applications. Similarly, Hosseinzadeh et al. (2021) applied SVM to truck-involved crashes in Iran, outperforming a Random Parameter Logit model and identifying driver fatigue and leftward deviation as major contributors to severe truck accidents.

Handling imbalanced datasets remains a critical challenge in accident severity prediction. Fiorentini and Losa (2020) employed Random Undersampling of the Majority Class (RUMC) to improve predictive performance across four machine learning models: Random Tree, k-Nearest Neighbor, Logistic Regression, and Random Forest. Their findings indicated that models incorporating RUMC achieved superior accuracy, with RUMC-Logistic Regression performing best (62.53%).

In contrast, Ospina-Mateus et al. (2021) proposed a hybrid approach that integrated association rule mining within a multi-objective genetic algorithm (NSGA-II) to enhance motorcycle crash injury prediction. Their methodology improved accuracy, precision, and recall by 20–21% over traditional machine learning models, demonstrating the potential of hybrid techniques in injury severity analysis. Çeven and Albayrak (2024) proposed decision tree-based models to predict traffic accident severity using a dataset of 23,074 urban traffic accidents in Kayseri, Türkiye, from 2013 to 2021. Ensemble learning methods, including Random Forest, AdaBoost, and Multi-Layer Perceptron (MLP), were employed to analyze 15 selected features. The F1 scores for these models ranged from 88.95% to 91.72%, with Random Forest achieving the highest performance. The study also identified "driver fault" as the most influential variable in determining accident severity, suggesting that targeted safety measures focused on driver behavior could reduce traffic accidents in urban areas.

Additionally, Tang et al. (2025) advanced accident severity prediction by integrating Multi-Layer Perceptron (MLP) with Decision Tree, XGBoost, and Random Forest models. Their "MLP + Random Forest" fusion model achieved an accuracy of 94%, showcasing its capability in real-time accident prediction and traffic safety management.

This section highlights the ongoing evolution of machine learning techniques in traffic accident severity prediction. While XGBoost and ensemble learning methods consistently demonstrate high performance, advancements in hybrid modeling and data balancing techniques continue to refine predictive accuracy and reliability. These insights contribute to the broader effort of leveraging machine learning for improved traffic safety management and accident mitigation strategies.

3.3 Factor-specific studies: influence of road, environmental, and vehicle-related factors

This category focuses on studies examining the influence of specific variables, such as road conditions, weather, and vehicle type, on crash severity. These studies provide valuable insights into how various factors contribute to the outcomes of traffic accidents, thereby informing strategies for improving traffic safety. Several studies have emphasized the role of road conditions and environmental factors, such as weather, in influencing traffic accident severity. For instance, Goswamy et al. (2023) examined the effectiveness of Rectangular Rapid Flashing Beacons (RRFB) at pedestrian crossings and found that the installation of RRFBs significantly reduced the severity of nighttime and pedestrian-related crashes. Their study demonstrated the positive impact of RRFBs on safety, particularly in nighttime conditions, with the XGBoost model yielding an accuracy of 97%.

Yang et al. (2025) analyzed how road surface conditions affect injury severity at highway-rail grade crossings. Their study, which used LightGBM and Shapley Additive Explanations (SHAP) values, identified several critical factors influencing injury severity, including train speed, driver age, and road surface type. The study recommended speed limits for both road vehicles and trains and the removal of surface obstacles to enhance safety at these crossings, particularly under adverse weather conditions.

Studies also explore how vehicle type and driver behavior, such as safety belt usage and psychophysical conditions, influence crash severity. Pillajo-Quijia et al. (2020) focused on run-off-road and rollover crashes involving light trucks and vans in Spain. They identified critical factors such as safety belt usage, alcohol consumption, drug use, and sleep deprivation as significant contributors to injury severity, highlighting the need for interventions targeting these behaviors.

In addition to examining specific factors, some studies have focused on addressing challenges such as imbalanced crash severity data and improving predictive model performance. Jiang et al. (2022) tackled the issue of modeling highly imbalanced crash severity data, where most crashes are classified as Property-Damage-Only (PDO). The study used AdaBoost and Gradient Boosting ensemble methods, achieving superior performance over other techniques. Key factors influencing crash severity included road characteristics, accident type, and seat belt usage, providing important insights into how these factors should be incorporated into predictive models. Mondal et al. (2020) compared the performance of Random Forest (RF) and Bayesian Additive Regression Trees (BART) in predicting weather-related crash severity. The study demonstrated that RF outperformed BART in terms of predictive accuracy. However, the study was limited by a lack of detailed information regarding the model implementation process and did not address the broader generalizability of the findings, which are important considerations for the application of machine learning techniques in real-world traffic safety scenarios.

Abdullah and Sipos (2024) focused on the spatial and temporal aspects of traffic accidents in Budapest, using decision trees and spatial heatmaps to identify high-risk locations. They found that the city center was the most accident-prone area, with crash severity influenced by time of day, lane configuration, and the day of the week. Their findings emphasize the importance of targeting high-risk locations and implementing data-driven mitigation strategies.

These studies underscore the complex interactions between road, environmental, and vehicle-related factors in determining crash severity. By utilizing machine learning models, these studies provide valuable insights into improving the prediction and mitigation of traffic accidents.

3.4 Crash type-specific prediction models

This category of studies focuses on the application of machine learning techniques tailored to specific crash scenarios, such as two-wheeler accidents, hazardous material transport incidents, and highway crashes. These models aim to provide insights into the unique factors influencing crash severity within each context, allowing for more precise and context-sensitive predictions. Gan and Weng (2020) explored injury severity in highway crashes by distinguishing between incidents involving traffic hazards and those that did not. The study analyzed 4,841 highway crashes across four provinces in China (Hubei, Guizhou, Chongqing, and Guangdong), occurring between January 2017 and April 2019. The authors employed Multinomial Logit and Random Forest models to predict injury severity. The performance of the models was evaluated using precision, recall, and F1 score metrics, with the Random Forest model demonstrating superior accuracy compared to the Multinomial Logit model. The study identified that time of day was the most significant factor in crashes involving traffic hazards, while vehicle type was the primary determinant in crashes without traffic hazards. These findings highlight the importance of tailoring safety measures to specific crash scenarios based on influential factors such as time and vehicle characteristics.

Al-Moqri et al. (2020) compared the performance of a Multinomial Logit model with five machine learning classifiers – Naïve Bayes, J48 Decision Tree, Random Forest, Support Vector Machine, and Multilayer Perceptron – to model injury severity in road traffic crashes. The study utilized data from crashes that occurred over a two-month period (August-October 2015), involving 156 injured individuals. To address data imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was employed. The classifiers were evaluated using a variety of performance metrics, including accuracy, precision, recall, F-measure, and Kappa statistics. Among all models, Random Forest achieved the highest performance, with an accuracy of 94.84%. The study identified several key factors influencing injury severity, including road type, the number of injured individuals, crash type, road users, transport mode to the emergency department, and crash action. These results underscore the importance of using machine learning to capture the complexities of crash scenarios and to prioritize safety measures based on influential variables.

Panicker and Ramadurai (2022) developed a machine learning model to predict injury severity in motorized two-wheeler (TW) crashes at mid-block road sections in Chennai, Tamil Nadu. The study utilized 7,654 TW crash cases from 2016 to 2018 and compared the performance of Random Forest (RF), Conditional Inference Forest (Cforest), and Ordered Probit (OP) models. The results indicated that Cforest outperformed both RF and OP models, particularly in capturing interaction effects that the other models failed to identify. Significant factors influencing injury severity in two-wheeler crashes included the type of colliding vehicle, collision type, driver age, and road visibility. This study highlights the importance of considering specific environmental and demographic factors when predicting injury severity in two-wheeler accidents, providing a more nuanced understanding of the risks associated with these types of crashes.

Rana et al. (2023) examined the factors influencing crashes at highway-railroad grade crossings (HRGCs) in Canada, focusing on identifying major causal factors and the severity of associated casualties. The study utilized ensemble machine learning algorithms to analyze HRGC crash data spanning from 2001 to 2022. Spatial autocorrelation and hotspot analysis, conducted using ArcGIS, revealed key HRGC crash clusters around major Canadian cities. Among the machine learning models, XGBoost achieved the highest mean accuracy (0.90), followed by Random Forest (0.87) and AdaBoost (0.82). The study's findings provided valuable insights into HRGC crash hotspots and causal factors, offering a foundation for targeted safety improvements at these critical locations. This research underscores the significance of using machine learning to identify and mitigate the risks associated with HRGC crashes, with implications for both infrastructure and policy interventions.

These studies emphasize the importance of using machine learning models tailored to specific crash types to improve the accuracy of severity predictions. By accounting for factors such as vehicle type, crash location, and environmental conditions, these models enable more precise safety interventions that can be targeted to specific high-risk scenarios.

3.5 Hybrid and ensemble approaches for enhanced prediction

The studies in this category investigate the application of hybrid and ensemble modeling techniques aimed at enhancing the accuracy and reliability of accident severity predictions. By combining different models or methods, these approaches seek to capitalize on the strengths of individual algorithms to achieve improved performance across diverse accident scenarios. Yan and Shen (2022) explored the trade-off between prediction accuracy and model interpretability in traffic accident severity prediction using a Bayesian Optimized Random Forest (BO-RF). Their findings demonstrated superior performance, particularly in recall, F1 score, and AUC, although limitations such as bias due to imbalanced data and a narrow range of input features were noted. These challenges highlight the need for further model refinement to address biases and expand feature sets for more comprehensive predictions. Similarly, Tang et al. (2020) reviewed various statistical and machine learning methods for predicting road incident clearance times, comparing models like Random Forest (RF) and Random Parameters Hazard-Based Duration (RPHD) with K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Back Propagation Neural Network (BPNN). The study found that RF and RPHD were particularly effective in data fitting and prediction, with heterogeneity models outperforming traditional machine learning techniques. Incident-specific factors such as incident type and lane closure were identified as significant predictors. Aldhari et al. (2023) focused on traffic accident severity prediction in Qassim Province, Saudi Arabia, using three classifiers: XGBoost, Random Forest, and Logistic Regression. The study found that XGBoost outperformed the other models in both multi-class and binary classification, with key predictors including road type, lighting conditions, and vehicle count, underscoring XGBoost's potential for injury severity prediction across various traffic contexts. Lastly, Peng et al. (2024) investigated the temporal evolution of factors influencing traffic accident severity during the COVID-19 pandemic, using a Random Forest model. While the study demonstrated how factors like traffic signals and weather conditions fluctuated in importance over time, its geographic limitation to Illinois and exclusion of driver- and vehicle-specific data underscored the need for broader data integration to fully capture the complex dynamics affecting traffic accident severity. Together, these studies highlight the critical role of hybrid and ensemble models in advancing traffic safety prediction, while also emphasizing the need for more inclusive data and refined methodologies to enhance model performance and real-world applicability.

4 Conclusion

The review of recent studies on machine learning techniques for traffic accident severity prediction underscores the growing importance of utilizing advanced algorithms to enhance prediction accuracy and model interpretability. Among the various models examined, Random Forest (RF) consistently emerged as a leading algorithm, achieving high accuracy and precision across diverse studies (e.g., Al-Moqri et al., 2020; Umer et al., 2020; Zhao et al., 2022). XGBoost also demonstrated notable performance, with studies such as Shen and Wei (2020) and Yang et al. (2022) highlighting its ability to effectively capture complex patterns in accident severity, especially when combined with other methods like Bayesian optimization or Bayesian networks. Additionally, ensemble approaches, including Voting Classifiers and Gradient Boosting Machines, have shown considerable promise in improving prediction outcomes, as evidenced by the findings of Jiang et al. (2022) and Panicker and Ramadurai (2022). The integration of interpretable models, such as Shapley Additive Explanations (SHAP), has facilitated a deeper understanding of the influential factors behind crash severity, including road conditions, vehicle type, and driver behavior (e.g., Guo et al., 2021; Zhao et al., 2022). Despite the success of these algorithms, certain challenges remain, particularly in handling imbalanced datasets, incorporating temporal variations, and ensuring model generalizability across different geographical contexts. Future research should focus on refining these models, incorporating a broader set of features, and improving their robustness in real-world applications to ensure more accurate and actionable traffic safety predictions.

Acknowledgement

The authors thank the Faculty of Transportation Engineering and Vehicle Engineering of the Budapest University of Technology and Economics, and the Department of Transport Technology and Economics for supporting this research.

References

Abdullah, P., Sipos, T. (2022) "Drivers' behavior and traffic accident analysis using decision tree method", Sustainability, 14(18), 11339. https://doi.org/10.3390/su141811339

Abdullah, P., Sipos, T. (2024) "Exploring the factors influencing traffic accidents: An analysis of black spots and decision tree for injury severity", Periodica Polytechnica Transportation Engineering, 52(1), pp. 33–39. https://doi.org/10.3311/PPtr.22392

Aldhari, I., Almoshaogeh, M., Jamal, A., Alharbi, F., Alinizzi, M., Haider, H. (2023) "Severity prediction of highway crashes in Saudi Arabia using machine learning techniques", Applied Sciences, 13(1), 233. https://doi.org/10.3390/app13010233

Al-Moqri, T., Haijun, X., Namahoro, J. P., Alfalahi, E. N., Alwesabi, I. (2020) "Exploiting machine learning algorithms for predicting crash injury severity in Yemen: hospital case study", Applied and Computational Mathematics, 9(5), pp. 155–164. https://doi.org/10.11648/j.acm.20200905.12

Arhin, S. A., Gatiba, A. (2020) "Predicting crash injury severity at unsignalized intersections using support vector machines and naïve Bayes classifiers", Transportation Safety and Environment, 2(2), pp. 120–132. https://doi.org/10.1093/tse/tdaa012

Bokaba, T., Doorsamy, W., Paul, B. S. (2022) "Comparative study of machine learning classifiers for modelling road traffic accidents", Applied Sciences, 12(2), 828. https://doi.org/10.3390/app12020828

Çeven, S., Albayrak, A. (2024) "Traffic accident severity prediction with ensemble learning methods", Computers and Electrical Engineering, 114, 109101. https://doi.org/10.1016/j.compeleceng.2024.109101

Chen, M.-M., Chen, M. C. (2020) "Modeling road accident severity with comparisons of logistic regression, decision tree and random forest", Information, 11(5), 270. https://doi.org/10.3390/info11050270

Cicek, E., Akin, M., Uysal, F., Topcu Aytas, R. (2023) "Comparison of traffic accident injury severity prediction models with explainable machine learning", Transportation Letters, 15(9), pp. 1043–1054. https://doi.org/10.1080/19427867.2023.2214758

Daoud, R., Vechione, M., Gurbuz, O., Sundaravadivel, P., Tian, C. (2025) "Comparison of machine learning models to predict nighttime crash severity: A case study in Tyler, Texas, USA", Vehicles, 7(1), 20. https://doi.org/10.3390/vehicles7010020

Elsevier "ScienceDirect", [online] Available at: https://www.sciencedirect.com [Accessed: 07 December 2024]

Elsevier "Scopus", [online] Available at: https://www.scopus.com [Accessed: 07 December 2024]

Fiorentini, N., Losa, M. (2020) "Handling imbalanced data in road crash severity prediction by machine learning algorithms", Infrastructures, 5(7), 61. https://doi.org/10.3390/infrastructures5070061

Gan, X., Weng, J. (2020) "Predicting crash injury severity for the highways involving traffic hazards and those involving no traffic hazards", In: CICTP 2020, Xi'an, China, pp. 4195–4206. ISBN 978078448293 https://doi.org/10.1061/9780784482933.360

Garrido, R., Bastos, A., de Almeida, A., Elvas, J. P. (2014) "Prediction of road accident severity using the ordered probit model", Transportation Research Procedia, 3, pp. 214–223. https://doi.org/10.1016/j.trpro.2014.10.107

Goswamy, A., Abdel-Aty, M., Islam, Z. (2023) "Factors affecting injury severity at pedestrian crossing locations with Rectangular RAPID Flashing Beacons (RRFB) using XGBoost and random parameters discrete outcome models", Accident Analysis and Prevention, 181, 106937. https://doi.org/10.1016/j.aap.2022.106937

Guo, M., Yuan, Z., Janson, B., Peng, Y., Yang, Y., Wang, W. (2021) "Older pedestrian traffic crashes severity analysis based on an emerging machine learning XGBoost", Sustainability, 13(2), 926. https://doi.org/10.3390/su13020926

Gururaj, H. L., Janhavi, V., Tanuja, U., Flamini, F., Soundarya, B. C., Ravi, V. (2022) "Predicting traffic accidents and their injury severities using machine learning techniques", International Journal of Transport Development and Integration, 6(4), pp. 363–377. https://doi.org/10.2495/TDI-V6-N4-363-377

Hee, L. V., Khamis, N., Noor, R. M., Abdul Karim, S. A., Puspitasari, P. (2024) "Predicting Fatality in Road Traffic Accidents: A Review on Techniques and Influential Factors", In: Abdul Karim, S. A. (ed.) Intelligent Systems Modeling and Simulation III: Artificial Intelligent, Machine Learning, Intelligent Functions and Cyber Security, Springer, Cham, pp. 343–367. ISBN 978-3-031-67317-7 https://doi.org/10.1007/978-3-031-67317-7_21

Hosseinzadeh, A., Moeinaddini, A., Ghasemzadeh, A. (2021) "Investigating factors affecting severity of large truck-involved crashes: Comparison of the SVM and random parameter logit model", Journal of safety research, 77, pp. 151–160. https://doi.org/10.1016/j.jsr.2021.02.012

IEEE "IEEE Xplore", [online] Available at: https://ieeexplore.ieee.org/Xplore/home.jsp [Accessed: 07 December 2024]

Infante, P., Jacinto, G., Afonso, A., Rego, L., Nogueira, V., Quaresma, P., Saias, J., Santos, D., Nogueira, P., Silva, M., Costa, R. P., Gois, P., Manuel, P. R. (2022) "Comparison of statistical and machine-learning models on road traffic accident severity classification", Computers, 11(5), 80. https://doi.org/10.3390/computers11050080

Jiang, L., Xie, Y., Wen, X., Ren, T. (2022) "Modeling highly imbalanced crash severity data by ensemble methods and global sensitivity analysis", Journal of Transportation Safety and Security, 14(4), pp. 562–584. https://doi.org/10.1080/19439962.2020.1796863

Jima, D., Sipos, T. (2022) "The impact of road geometric formation on traffic crash and its severity level", Sustainability, 14(14), 8475. https://doi.org/10.3390/su14148475

Mondal, A. R., Bhuiyan, M. A. E., Yang, F. (2020) "Advancement of weather-related crash prediction model using nonparametric machine learning algorithms", SN Applied Sciences, 2, 1372. https://doi.org/10.1007/s42452-020-03196-x

Obasi, I. C., Benson, C. (2023) "Evaluating the effectiveness of machine learning techniques in forecasting the severity of traffic accidents", Heliyon, 9(8), e18812. https://doi.org/10.1016/j.heliyon.2023.e18812

Ospina-Mateus, H., Quintana Jiménez, L. A., Lopez-Valdes, F. J., Berrio Garcia, S., Barrero, L. H., Sana, S. S. (2021) "Extraction of decision rules using genetic algorithms and simulated annealing for prediction of severity of traffic accidents by motorcyclists", Journal of Ambient Intelligence and Humanized Computing, 12(11), pp. 10051–10072. https://doi.org/10.1007/s12652-020-02759-5

Ötvös, V., Török, Á. (2024) "Measurement of accident risk and a case study from Hungary", Periodica Polytechnica Transportation Engineering, 52(2), pp. 159–165. https://doi.org/10.3311/PPtr.22731

Panicker, A. K., Ramadurai, G. (2022) "Injury severity prediction model for two-wheeler crashes at mid-block road sections", International Journal of Crashworthiness, 27(2), pp. 328–336. https://doi.org/10.1080/13588265.2020.1806644

Peng, Q., Bakkar, Y., Wu, L., Liu, W., Kou, R., Liu, K. (2024) "Transportation resilience under Covid-19 uncertainty: A traffic severity analysis", Transportation Research Part A: Policy and Practice, 179, 103947. https://doi.org/10.1016/j.tra.2023.103947

Pillajo-Quijia, G., Arenas-Ramírez, B., González-Fernández, C., Aparicio-Izquierdo, F. (2020) "Influential factors on injury severity for drivers of light trucks and vans with machine learning methods", Sustainability, 12(4), 1324. https://doi.org/10.3390/su12041324

Rana, P., Sattari, F., Lefsrud, L., Hendry, M. (2023) "Machine learning approach to enhance highway railroad grade crossing safety by analyzing crash data and identifying hotspot crash locations", Transportation Research Record, 2678(7), pp. 1055–1071. https://doi.org/10.1177/03611981231212162

Rifaat, S. M., Chin, H. C. (2007) "Accident severity analysis using ordered probit model", Journal of Advanced Transportation, 41(1), pp. 91–114. https://doi.org/10.1002/atr.5670410107

Shen, X., Wei, S. (2020) "Application of XGBoost for hazardous material road transport accident severity analysis", IEEE Access, 8, pp. 206806–206819. https://doi.org/10.1109/access.2020.3037922

Sipos, T., Afework Mekonnen, A., Szabó, Z. (2021) "Spatial econometric analysis of road traffic crashes", Sustainability, 13(5), 2492. https://doi.org/10.3390/su13052492

Sorum, N. G., Pal, D. (2024) "Identification of the best machine learning model for the prediction of driver injury severity", International Journal of Injury Control and Safety Promotion, 31(3), pp. 360–375. https://doi.org/10.1080/17457300.2024.2335478

Tang, J., Huang, Y., Liu, D., Xiong, L., Bu, R. (2025) "Research on traffic accident severity level prediction model based on improved machine learning", Systems, 13(1), 31.
https://doi.org/10.3390/systems13010031

Tang, J., Zheng, L., Han, C., Yin, W., Zhang, Y., Zou, Y., Huang, H. (2020) "Statistical and machine-learning methods for clearance time prediction of road incidents: A methodology review", Analytic Methods in Accident Research, 27, 100123.
https://doi.org/10.1016/j.amar.2020.100123

Török, Á. (2017) "Comparative analysis between the theories of road transport safety and emission", Transport, 32(2), pp. 192–197.
https://doi.org/10.3846/16484142.2015.1062798

Umer, M., Sadiq, S., Ishaq, A., Ullah, S., Saher, N., Madni, H. A. (2020) "Comparison analysis of tree based and ensembled regression algorithms for traffic accident severity prediction", [preprint] arXiv, (27 October 2020)
https://doi.org/10.48550/arXiv.2010.14921

Wang, C., Chen, F., Zhang, Y., Cheng, J. (2023) "Analysis of injury severity in rear-end crashes on an expressway involving different types of vehicles using random-parameters logit models with heterogeneity in means and variances", Transportation Letters, 15(7), pp. 742–753.
https://doi.org/10.1080/19427867.2022.2086760

Yan, M., Shen, Y. (2022) "Traffic accident severity prediction based on random forest", Sustainability, 14(3), 1729.
https://doi.org/10.3390/su14031729

Yang, Y., Wang, K., Yuan, Z., Liu, D. (2022) "Predicting freeway traffic crash severity using XGBoost-Bayesian network model with consideration of features interaction", Journal of Advanced Transportation, 2022(1), 4257865.
https://doi.org/10.1155/2022/4257865

Yang, Z., Zhang, C., Li, G., Xu, H. (2025) "Analysis of the impact of different road conditions on accident severity at highway-rail grade crossings based on explainable machine learning", Symmetry, 17(1), 147.
https://doi.org/10.3390/sym17010147

Zhao, G., Jiang, Y., Li, S., Tighe, S. (2022) "Exploring implicit relationships between pavement surface friction and vehicle crash severity using interpretable extreme gradient boosting method", Canadian Journal of Civil Engineering, 49(7), pp. 1206–1219.
https://doi.org/10.1139/cjce-2021-0337

Zou, Y., Lin, B., Yang, X., Wu, L., Muneeb Abid, M., Tang, J. (2021) "Application of the bayesian model averaging in analyzing freeway traffic incident clearance time for emergency management", Journal of Advanced Transportation, 2021(1), 6671983.
https://doi.org/10.1155/2021/6671983