Journal of Scientific Insight
E-ISSN: 3062-8571
DOI: https://doi.org/10.69930/jsi.v1i4.237
Review Article

Vol. 1 (4), 2024
Page: 170-179

Data-Driven Road Safety: A Machine Learning Framework Utilizing Open Traffic Data

Mohamed H. Abdelati [1,*], Al-Hussein Matar [1], Hilal A. Abdelwali [2], Ebram F.F. Mokbel [1], M. Rabie [1]

[1] Department of Automotive and Tractors, Faculty of Engineering, Minia University, Egypt
[2] Department of Automotive and Marine, College of Technological Studies, Kuwait

*Email (corresponding author): m.hilal@mu.edu.eg

Abstract
Road traffic accidents continue to be a problem across the world and according to statistics cause high mortality and economic losses. This research work conceptualizes an idea that will use open traffic data and machine learning models to forecast accidents on roads in order to promote road safety. Based on the presented literature review, the framework incorporates a step-by-step procedure to analyze risk factors for targeted safety interventions, including data pre-processing and feature selection, application of a chosen model for high-risk zones identification, and improving the result by altering related factors. The findings show the applicability of open data and predictive analysis in traffic safety matters, with special emphasis on temporal, spatial, and environmental features. Resources allocation, urban traffic control, and monitoring are cases used to illustrate the framework's applicability. Although this is a conceptual model, the challenges, such as data quality, data privacy issues, and practical issues with implementation, are also included in the framework, along with suggestions for future research, such as the use of stream data and improved modeling techniques. This investigation contributes to the literature as a robust theoretical model from which practical solutions for road traffic safety interventions can be derived to reduce and ultimately eliminate traffic accidents and fatalities worldwide.

Keywords: *AI-Driven traffic management, traffic flow optimization, smart city traffic solutions, real-time traffic prediction, sustainable urban mobility*

1. Introduction

Road traffic injuries present themselves as modifiable causes of morbidity and mortality as well as sources of economic loss in the contemporary world(Abdelati, 2024; Gabr, Shoaeb, & El-Badawy, 2018; Nagy & Simon, 2018). Nevertheless, due to the intensification of innovative vehicle systems and infrastructural developments, the figures of causality and injuries still have not come up in accordance with the innovations, particularly in the congested and the conflict-laden urban areas(Berkowicz, Winther, & Ketzel, 2006; Mohan, Khayesi, Tiwari, & Nafukho, 2006). Solving this problem demands that some action plans are taken that are not necessarily reactive in nature(El-Wahab, Rabie, Abdelati, Khalil, & Abdelgawwad, 2021; H Abdelati & Abdelhafeez, 2023; Jereb, Stopka, & Skrúcaný, 2021).

While current traffic safety measures are largely reactive, addressing accidents after they occur, there is an increasing focus on proactive measures. Proactive strategies are designed

Received: 22 November 2024 / Accepted: 23 December 2024

to prevent accidents before they happen by predicting and managing traffic flow, adjusting signal timings, and implementing real-time monitoring systems.

Proactive approaches are proven to reduce both the occurrence of accidents and their severity. For example, technologies such as traffic prediction models and automated traffic management systems can anticipate congestion and adjust traffic signals to avoid accidents before they occur.

As traffic datasets are becoming more open designed and accessible, a new field of study has been created(Chen, Lv, Li, & Wang, 2016; Rohunen, Markkula, Heikkila, & Heikkila, 2014). These datasets offer precise information on the road conditions, past accidents, local climate, and traffic patterns, which remain unexplored for analytical value(Abdelati, 2023; Gregurić, Vujić, Alexopoulos, & Miletić, 2020). The obtained data can be combined with machine learning algorithms to foresee high-risk situations and help the authorities put in force safeguards in advance(Abdelati, Khalil, Abdelgawwad, & Rabie, 2020; Alardhi, Abdelwali, Khalfan, & Abdelati).

The novelty of this study lies in its innovative approach of leveraging open traffic data in conjunction with machine learning techniques to proactively predict and prevent road traffic accidents. Unlike traditional reactive safety measures, which address accidents post-occurrence, this study focuses on predictive modeling and real-time intervention strategies. By identifying high-risk situations before they lead to accidents, this approach aims to significantly reduce both the frequency and severity of traffic-related injuries, offering a transformative shift in how road safety can be managed.

By investigating proactive strategies, this study aims to contribute to enhancing road safety, reducing traffic accidents, and ultimately saving lives and resources. In light of this, this paper aims to investigate the use of open traffic data and machine learning as a step towards increasing road safety. Thus, through a systematic review of the literature, it is hoped that the best methodologies can be discerned and existing weaknesses in practice can be highlighted. This study outlines a conceptual model for accident prediction that is potentially useful for future empirical and application-oriented research. The goal of this work is to act as a reference point for both academics and professionals in search of relevant and effective traffic safety interventions that are grounded in data.

2. Literature Review

Open data and machine learning are now combined to help reduce traffic accidents, providing an opportunity to learn and predict concerning road safety(Cho, Mitsuya, & Kato, 2000). Online databases containing information on historical accidents, traffic flow and environmental factors are obtained from government databases, global project data and municipal websites(Brinkhoff, 2003; D'Alconzo, Drago, Morichetta, Mellia, & Casas, 2019). Such datasets are the starting point of machine learning models, which facilitate realistic assessment and recognition of patterns.

These aspects have been proven to hold a great chance that is brought by machine learning techniques in the prediction of traffic accidents(Bachechi, Po, & Rollo, 2022; Chong, Abraham, & Paprzycki, 2005). Demographic data together with Random forest, gradient boosting, and Neural networks have been used to predict areas that are most prone to accidents, examine the causes, and recommend control measures. Research has also delved into the use of the concept of voting using several algorithms to get a single enhanced predictive accuracy(Abdelati, Abd-El-Tawwab, Ellimony, & Rabie, 2024; AlMamlook, Kwayu, Alkasisbeh, & Frefer, 2019). For example, the neural network algorithm is most efficient in dealing with nonlinear dependencies between broken factors such as weather, road condition, and accident severity The decision tree mainly deals with categorical variables such as time of day or driver behavior.

However, a few limitations exist that limit the application of machine learning for improving traffic safety(Bokaba, Doorsamy, & Paul, 2022; Najafi Moghaddam Gilani, Hosseinian, Ghasedi, & Nikookar, 2021). There are hundreds and thousands of research works carried out on traffic systems, whereby many use databases that do not have dynamic properties of the system, thus limiting the practical use of the findings. Second, the obtained open data may not have the same quality or contain some fields at all, which causes various biases distorting the results of the forecast(Adewopo et al., 2023; Irwan & Mursyid, 2024; Shaygan, Meese, Li, Zhao, & Nejad, 2022). The absence of specific techniques for data preparation, input variable selection and model assessment makes it difficult to compare results from different related research works(Siswanto, Syaban, & Hariani, 2023).

These challenges show that though commendable progress has been made, a lot needs to be done, and this is where coming up with extraordinary solutions comes in(Banerjee et al., 2022; Dogru & Subasi, 2012). Challenges include improving the architecture for the operation of real-time data feeds, increasing the data quality for the generation of data-driven algorithms, and adopting standard structures to the machine learning operations. The closeness of these gaps is central in the progression of the application of prediction models in the modulation of traffic accidents and road safety(Halim, Kalsoom, Bashir, & Abbas, 2016).

3. Proposed Framework

3.1. Conceptual Workflow

In this paper, a methodology framework is developed, which provides a structured approach of applying open traffic data and machine learning for traffic accident prediction. The tasks within this work flow start with the collection of this data from public transport data providers such as accident history, traffic density, weather condition, and the roads' characteristics. The data is preprocessed to remove outliers and missing values, wash and normalization, and to handle such aspects as imbalances.

Next is the feature selection process attention is dedicated to important predictors for example temporal features (time of day), space features (for example proximity to crossings), and environmental aspects (for example weather conditions). Then machine learning models are trained on the preprocessed data using Interpretability using Random Forests and capturing of Nonlinear relationships using Neural Networks. Last but not least, the framework focuses on the visualization of the prediction results using geographical information system tools so that policymakers can pinpoint the area's most vulnerable to the problem being addressed in this paper. Figure 1 Workflow illustrating the proposed framework from data input and preprocessing to accident prediction and real-time outputs.

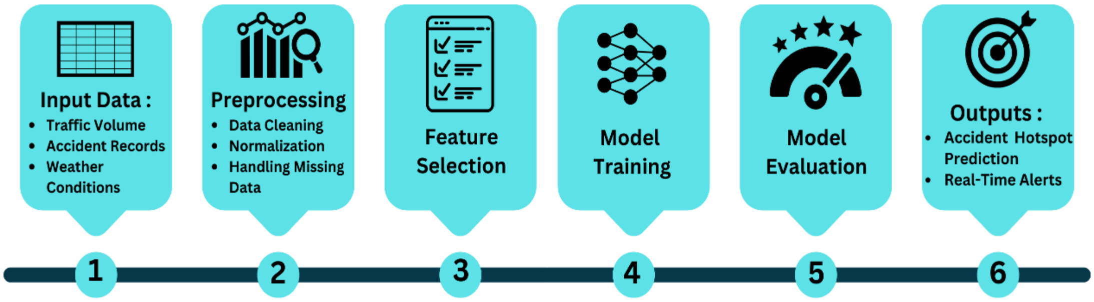

Figure 1. Proposed framework workflow from data input to accident prediction

3.2. Key Features and Predictors

The proposed framework helps in giving a solution to prioritize the features, which are vital for accident prediction. The temporal features including the speed of flow during peak traffic hours as well as holiday influences are useful in revealing the accident-prone periods. Traffic characteristics of road networks, signal sightlines, and existing congestion, are considered vital in determining risky areas. Which in addition to lidar and radar data includes rainfall, visibility, and other characteristics of the road surface. This makes the interaction of these variables deemed to reveal these relationships assuredly, making the framework to consider real-life complexities. Table (1) highlights the main categories of features, their associated predictors, the required data sources, and their relationship to accident severity.

Table 1. Key features and predictors in traffic accident prediction

Category	Predictors	Data Required/Sources	Relation to Accident Severity
Temporal	Time of Day, Day of the Week, Seasonal Trends	Traffic system logs, police reports.	Primarily related to accident probability during peak hours.
Spatial	Road Geometry, Proximity to Intersections, Speed Limits	OpenStreetMap data, infrastructure records.	Impacts both accident probability and severity.
Environmental	Weather Conditions (e.g., rain, fog), Lighting Conditions	Meteorological data, surveillance records.	Strongly influences accident severity.
Traffic Dynamics	Traffic Volume, Congestion Levels, Vehicle Types	Traffic sensors, intelligent monitoring systems.	Mainly linked to accident probability rather than severity.
Driver Behavior	Speeding, Distraction Events, Fatigue Reports	Police reports, telematics devices.	Highly impacts both accident probability and severity.

3.3. Suggested Modeling Techniques

The framework promotes the use of multiple models to achieve higher prediction accuracy and model update flexibility. Random Forests are emphasized with regard to its applicability with mixed data types and for analyzing feature importance. Neural networks are advised since they are ideal for solving high-degree non-interactive models specially with the help of big data. Other techniques like Gradient Boosting put together the results of multiple classifiers since combining the results of several methods gives higher accuracy. The choice of algorithms is based on the characteristics of the data and the level of interpretability of the decision versus the degree of accuracy.

This framework provides the theoretical background for future research, providing scalability and flexibility for works based on open traffic data for predictive road safety applications.

4. Practical Applications

4.1. Urban Traffic Management

The proposed framework has a vast potential to help in managing urban traffic concerning traffic accident occurrences. Application of mathematic models that would allow the early detection of provinces that need extra security, for example, intersections, which always result in accidents, or zones, which have intense traffic flux, every day that will allow to prevent accidents. For instance, the use of extra traffic police during rush hour or in zones prone to accidents, or the determination of continually changing traffic light timings can minimize the probability of accidents. In addition, it can help with the development of better road design, therefore, the minimization of collisions.

4.2. Resource Allocation and Policy Development

Using the predictions of the areas where accidents are more likely to happen, authorities can make improvements better. First aid responders and other medical facilities can be placed near high-risk areas since it will take less time to get to the scene. The insights can also inform investment in road infrastructure where improvements are needed by identifying these areas. Moreover, case analyses that lead to recommendations for policy measures like speed control or the implementation of enhanced safety features in automobiles can be derived from the analyses of predictive outcomes.

4.3. Real-Time Monitoring and Alerts

Since the presented framework is developed at the moment of conception, its principles can be applied to work in real-time. Real-time risk analysis for traffic management can be derived from the IoT systems' devices; for instance, the traffic cameras and vehicle sensors. This could make real-time function and make use of indicators that could notify the drivers of likely dangers in the road including unfavorable weather conditions or traffic jams. Templates of this nature do not only act to improve the safety of the individual driver but also to address system objectives of traffic flow control with regard to avoiding multiple collision instances.

The specific recommendations derived from the proposed framework help to stress the value and versatility of the approach to both short-term and long-term traffic safety problems. It links academic research to practice, creating a framework for utilizing data analyses for the relevant practical purposes of accident avoidance and optimization of other resources.

5. Ethical and Practical Challenges

5.1. Data Privacy and Ethical Concerns

Open traffic data as the input for the predictive models brings several ethical issues into light especially right to privacy. Unlike other open datasets that may anonymize individual data points, aggregating such datasets with actual real-time data feeds, say IoT sensors or vehicle telematics data may expose extremely sensitive data. For example, anonymization of data, data protection acts, the general data protection regulation act etc. have to be met appropriately. Further on, original biases in the data sets—lack of presence of specific regions or minimized involvement of specific types of people—can result in inequality of protective measures which disadvantages the excluded community. These biases can only be addressed through consistent assessment of collection and modeling processes and alterations to remove biases.

5.2. Quality and Accessibility of Open Data

The usefulness of the proposed framework greatly depends on the quality and coverage of the input datasets. Big data is frequently unstructured and contains gaps as well as regional differences that can negatively affect the reliability of the forecast. However, several governments and organizations have adopted open data programs, availability and access to such datasets are not equitably distributed across the world hence the framework is somewhat relevant in these regions. These are the challenges of open data which can be solved partially by adopting standard data formats and increasing the number of participants in the projects.

5.3. Practical Barriers to Implementation

Applying the structure that has emerged for the proposed framework to concrete solutions is challenged by considerable practical hurdles. These are, for instance, the costs of the computational infrastructure necessary for large-scale data analysis and the costs of linking predictive systems into current Traffic Management Systems (TMS). Besides, there is always resistance from the stakeholders for example the local governments might prove to be reluctant in implementing change such as adopting complex technologies. Good examples, relevant lower cost, and safety claims are critical for enhancing adoption through the engagement of key stakeholders.

5.4. Balancing Interpretability and Complexity

Although newly developing sophisticated methods like the neural networks provide high accurate predictions, the problem of non-interpretableness arises in the context of traffic management and, particularly, traffic safety applications. Authors emphasize that these models should be understandable and give adequate recommendations to policymakers and practitioners. This tension in terms of the added difficulty involved in algorithms, while at the same time being required to be explainable, has not been properly addressed yet though work is being done through the use of explainable AI (XAI).

It is for these reasons important to address both the ethical as well as practical considerations to facilitate the efficacy of the aforementioned framework in the manner intended – by delivering impact, primarily, and secondly, by not compromising fairness, transparency and widening applicability. Therefore, by addressing these questions in advance, the presented framework can become a strong instrument for the further promotion of traffic safety activities globally.

6. Limitations and Recommendations

6.1. Conceptual Nature of the Framework

The proposed framework is in their nature more of a conceptual nature as it is built on theoretical assumptions and evidence taken from the literature not experimentation or application of data. Even though this approach reveals the possibility of utilizing open data and machine learning to enhance traffic safety, no proof of its efficiency has been established in practical application. The lack of field trials or simulations to validate some of the concepts limits its practical use in real, dynamic, and complex traffic systems.

6.2. Dataset Dependency

The fundamental weakness of the framework is the source of data, more so, where such data is sourced from open datasets that are characterized by variability in data quality, completeness, and regional variation. There are few issues that one may encounter while implementing this model, some regions may not have detailed data about the accidents or may not have access to raw traffic data. In addition, due to continually changing traffic conditions, it becomes nearly impossible to use static datasets to capture them, thus the importance of real-time data integration.

The use of open traffic datasets presents certain limitations that can affect the accuracy of predictive models. These datasets may suffer from issues such as incomplete or outdated information, biases towards specific geographic areas, and missing real-time data. Such deficiencies can lead to inaccuracies in predictions and may impact the generalization of results across different regions or traffic conditions. Moreover, these limitations may also result in predictive models being skewed, with some variables underrepresented or not accounted for, which can diminish the overall performance of the framework. Addressing these data quality concerns is crucial to improving the robustness and reliability of predictive modeling efforts.

6.3. Recommendations for Future Research

To enhance the utility and robustness of the framework, future research should focus on:

- Real-Time Data Integration: Engaging, IoT-connected objects and sensors, first- and third-party visual feeds to enhance the predictive models and their speed.
- Global Validation: Additional testing of the framework for scalability to different zones which have different traffic flow rates and patterns.
- Advanced Modeling Techniques: Introducing new AI methods, for example, reinforcement learning and federated learning, to define new approaches to overcome the weaknesses of TI used in traditional machine learning algorithms.
- Standardization Efforts: Collation of protocols for data acquisition, cleaning, and transformation, to ensure that the input data are harmonized to those routinely employed within investigations.

6.4. Addressing Broader Challenges

The proposed framework needs to expand their view of various contextual issues like the fair distribution of predictive instruments, as well as the compatibility of the framework with the policy and decision-making systems. Such interdisciplinary cooperation with urban planner and policymakers can go a long way in reducing the research practice gap. By noting the framework's shortcomings and offering specific suggestions on how to move forward, this paper offers a framework that will help open up traffic safety research and contribute to the development of further empirical and applied work.

Conclusions

This study presents a novel approach to traffic accident prediction by integrating open traffic data with machine learning techniques. By employing predictive modeling, the proposed framework enables the proactive identification of high-risk areas, paving the way for targeted safety interventions. The findings emphasize the importance of using real-time, open data to inform decision-making in traffic management systems. While the framework offers promising potential, its theoretical basis calls for empirical validation through real-world application to assess its efficacy.

The challenges related to data quality, privacy concerns, and implementation barriers must be addressed to maximize the framework's practical utility. However, the integration of open data and machine learning holds significant promise for transforming road safety management globally. This work lays the foundation for future research, fostering collaboration between researchers, policymakers, and industry experts to create data-driven, proactive solutions for reducing traffic accidents and improving public safety.

References

Adewopo, V., Elsayed, N., Elsayed, Z., Ozer, M., Wangia-Anderson, V., & Abdelgawad, A. (2023). Ai on the road: A comprehensive analysis of traffic accidents and accident detection system in smart cities. *arXiv preprint arXiv:2307.12128*.

Abdelati, M. H. (2023). Improving Solution Quality In Transportation Problems: A Novel Algorithm For Efficient Resource Allocation. *Agpe The Royal Gondwana Research Journal Of History, Science, Economic, Political And Social Science*, *4*(8), 1-10.

AlMamlook, R. E., Kwayu, K. M., Alkasisbeh, M. R., & Frefer, A. A. (2019). *Comparison of machine learning algorithms for predicting traffic accident severity.* Paper presented at the 2019 IEEE Jordan international joint conference on electrical engineering and information technology (JEEIT).

Abdelati, M. H. (2024). Smart Traffic Management for Sustainable Development in Cities: Enhancing Safety and Efficiency. *International Journal of Advanced Engineering and Business Sciences*, *5*(1).

Alardhi, M., Abdelwali, H. A., Khalfan, A. M., & Abdelati, M. H. Using the Minimize Distance Method to Find the Best Compromise Solution of Multi-objective Transportation Problem with Case Study.

Abdelati, M. H., Abd-El-Tawwab, A. M., Ellimony, E. E. M., & Rabie, M. (2024). " Efficient and Versatile Methodology for Solving Solid Transportation Problem Using Excel Solver: A Comparative Study with LINGO Code. *Industrial Engineering, 3*(1).

Abdelati, M. H., Khalil, M. I., Abdelgawwad, K., & Rabie, M. (2020). Alternative Algorithms For Solving Classical Transportation Problems. *Journal of Advanced Engineering Trends, 39*(1), 13-24.

Bachechi, C., Po, L., & Rollo, F. (2022). Big data analytics and visualization in traffic monitoring. *Big Data Research, 27*, 100292.

Banerjee, K., Bali, V., Sharma, A., Aggarwal, D., Yadav, A., Shukla, A., & Srivastav, P. (2022). *Traffic accident risk prediction using machine learning.* Paper presented at the 2022 International Mobile and Embedded Technology Conference (MECON).

Berkowicz, R., Winther, M., & Ketzel, M. (2006). Traffic pollution modelling and emission data. *Environmental Modelling & Software, 21*(4), 454-460.

Bokaba, T., Doorsamy, W., & Paul, B. S. (2022). Comparative study of machine learning classifiers for modelling road traffic accidents. *Applied Sciences, 12*(2), 828.

Brinkhoff, T. (2003). Generating traffic data. *IEEE Data Eng. Bull., 26*(2), 19-25.

Chen, Y.-y., Lv, Y., Li, Z., & Wang, F.-Y. (2016). *Long short-term memory model for traffic congestion prediction with online open data.* Paper presented at the 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC).

Cho, K., Mitsuya, K., & Kato, A. (2000). *Traffic data repository at the {WIDE} project.* Paper presented at the 2000 USENIX Annual Technical Conference (USENIX ATC 00).

Chong, M., Abraham, A., & Paprzycki, M. (2005). Traffic accident analysis using machine learning paradigms. *Informatica, 29*(1).

D'Alconzo, A., Drago, I., Morichetta, A., Mellia, M., & Casas, P. (2019). A survey on big data for network traffic monitoring and analysis. *IEEE Transactions on Network and Service Management, 16*(3), 800-813.

Dogru, N., & Subasi, A. (2012). *Traffic accident detection by using machine learning methods.* Paper presented at the Third international symposium on sustainable development (ISSD'12).

El-Wahab, A., Rabie, M., Abdelati, M. H., Khalil, M. I., & Abdelgawwad, K. (2021). A case study of reducing the total wasted time for the bus movement of Public Transportation Authority in Cairo (CTA). *SVU-International Journal of Engineering Sciences and Applications, 2*(2), 82-87.

Gabr, A., Shoaeb, A., & El-Badawy, S. (2018). Economic Impact Of Urban Traffic Congestion On The Main Routes In Mansoura City, Egypt. *International Journal for Traffic & Transport Engineering, 8*(2).

Gregurić, M., Vujić, M., Alexopoulos, C., & Miletić, M. (2020). Application of deep reinforcement learning in traffic signal control: An overview and impact of open traffic data. *Applied Sciences, 10*(11), 4011.

H Abdelati, M., & Abdelhafeez, A. (2023). An Integrated Approach for Promoting Sustainable Transportation: A Case Study of Fuel Consumption and Carbon Dioxide Emissions Reduction in the Cairo Transport Authority. *International Journal of Advanced Engineering and Business Sciences, 4*(3).

Halim, Z., Kalsoom, R., Bashir, S., & Abbas, G. (2016). Artificial intelligence techniques for driving safety and vehicle crash prediction. *Artificial Intelligence Review, 46*, 351-387.

Irwan, M., & Mursyid, M. (2024). AI-Driven Traffic Accidents: A Comparative Legal Study. *Artes Libres Law and Social Journal, 1*(1), 1-20.

Jereb, B., Stopka, O., & Skrúcaný, T. (2021). Methodology for estimating the effect of traffic flow management on fuel consumption and CO2 production: A case study of Celje, Slovenia. *Energies*, *14*(6), 1673.

Mohan, D., Khayesi, M., Tiwari, G., & Nafukho, F. M. (2006). *Road traffic injury prevention training manual*: World Health Organization.

Nagy, A. M., & Simon, V. (2018). Survey on traffic prediction in smart cities. *Pervasive and Mobile Computing, 50*, 148-163.

Najafi Moghaddam Gilani, V., Hosseinian, S. M., Ghasedi, M., & Nikookar, M. (2021). Data-Driven Urban Traffic Accident Analysis and Prediction Using Logit and Machine Learning-Based Pattern Recognition Models. *Mathematical problems in engineering, 2021*(1), 9974219.

Rohunen, A., Markkula, J., Heikkila, M., & Heikkila, J. (2014). Open traffic data for future service innovation: Addressing the privacy challenges of driving data. *Journal of theoretical and applied electronic commerce research, 9*(3), 71-89.

Shaygan, M., Meese, C., Li, W., Zhao, X. G., & Nejad, M. (2022). Traffic prediction using artificial intelligence: Review of recent advances and emerging opportunities. *Transportation research part C: emerging technologies*, *145*, 103921.

Siswanto, J., Syaban, A. S. N., & Hariani, H. (2023). Artificial intelligence in road traffic accident prediction. *Jambura Journal of Informatics, 5*(2), 77-90.