ACTA TECHNICA JAURINENSIS
Vol. XX, No. Y, pp. ZZ-ZZ, 2026
10.14513/actatechjaur.00951

Research Article

Forecasting the number of road accidents in Poland and Malaysia using neural networks

Piotr Gorzelańczyk[1,*], Jen Sim Ho[2]

[1]Department of Transport. Stanislaw Staszic State University of Applied Sciences in Pila, Podchorazych 10 Street, 64-920 Pila, Poland
[2] Malaysian Institute of Road Safety Research, Lot 125-135, Jalan TKS1, Taman Kajang Sentral, 43000 Kajang, Selangor, Malaysia
*e-mail: piotr.gorzelanczyk@ans.pila.pl

Submitted: 19/12/2025 Revised: 07/01/2026 Accepted: 20/01/2026 Published online: 05/02/2026

Abstract: Road traffic accidents remain a significant concern globally, including in Poland and Malaysia, despite an overall downward trend observed in recent years. While the COVID-19 pandemic has undeniably influenced accident rates, the figures still indicate a pressing need for proactive measures to further reduce their occurrence. This study aims to forecast the number of road traffic accidents in Malaysia and Poland between 2024 and 2030 using neural network models. Historical annual accident data from official national sources were analysed and processed using multilayer perceptron (MLP) neural networks. Two training-validation-testing data splits (70-15-15 and 80-10-10) were employed to assess model performance. The results indicate a likely stabilization in accident numbers in Poland, while Malaysia shows a post-pandemic increase. Despite the limitations of using aggregated data, the models achieved low prediction errors, with MAPE as low as 2.14% in the best configurations. These forecasts can inform evidence-based road safety policies in both countries. The results indicate that the number of road traffic accidents is expected to reach a stable level in the coming years. This outcome appears to be shaped by several key dynamics, including the ongoing development of transport infrastructure—particularly the addition of new highways and express routes—as well as the persistent rise in vehicle ownership. However, it should be emphasized that the reliability of these projections is constrained by the limitations inherent in the random sampling of datasets applied during the model's training, testing, and validation phases.

Keywords: traffic accident; pandemic; forecasting; neural networks

I. INTRODUCTION

Road traffic accidents remain a major challenge worldwide, carrying both severe public health consequences and substantial economic costs. The World Health Organization (WHO) estimates that each year approximately 1.3 million people lose their lives due to road traffic injuries, with these incidents consuming on average 3% of national GDP across the globe [1]. Alarmingly, crashes constitute the leading cause of mortality among individuals aged 5–29 [1]. In response, the United Nations General Assembly has adopted the ambitious goal of reducing global road traffic deaths and injuries by 50% by the year 2030.

Accurate evaluation of accident severity is essential for designing effective road safety policies. Identifying the determinants of crash outcomes provides the foundation for preventive measures and mitigation strategies [2–4]. For example, Yang et al. [5] introduced a Deep Neural Network (DNN) framework capable of predicting different levels of injury, fatality, and property loss, thereby supporting more precise and holistic assessments of crash severity.

The availability of high-quality data is a prerequisite for reliable accident modelling and forecasting. Conventionally, government bodies compile such information from official channels, including police records, insurance data, and hospital reports [6].

Advances in transportation technology have created additional sources of traffic-related information. Data streams from GPS-enabled vehicles, roadside microwave sensors, and automated license plate recognition systems provide continuous insights into vehicle movement, speed, and overall traffic flow [7–9]. Social media platforms also generate real-time traffic updates, though user-generated content may vary in reliability [10].

Robust accident analysis frequently benefits from integrating multiple datasets. By combining heterogeneous sources and reconciling data inconsistencies, researchers can achieve more accurate and comprehensive analytical outcomes [11].

Vilaca et al. [12] applied statistical techniques to explore the link between road users and accident severity, underlining the importance of stronger safety interventions and regulatory frameworks. Similarly, Bak et al. [13] employed multivariate statistical methods to examine risk factors associated with road users in a specific Polish region, using accident records as a surrogate for causation research.

Ultimately, the choice of data source must align with the research objective. Combining classical statistical approaches with real-world driving data and insights from intelligent transportation systems offers significant potential for enhancing accident prediction and mitigation strategies [14].

Time series models are widely used for accident frequency prediction [15, 16]. However, these models often exhibit residual autocorrelation [17] and may not effectively assess forecast accuracy based on past predictions. While some studies have employed methods like Holt-Winters exponential smoothing [19] and multi-seasonality models [18], these approaches may not effectively incorporate exogenous variables [20]. The controversial issue has also been considered by numerous researchers [21-30].

The number of road accidents per 10,000 inhabitants (NRA) serves as an important benchmark for assessing road safety performance. In 2023, Poland, with a population of 37.6 million, registered 20,936 crashes, corresponding to an NRA of 5.57. By comparison, Malaysia, home to 34.31 million residents, reported 598,635 accidents, which translates into an NRA of 174.78—substantially higher than the Polish rate.

$$NRA = \frac{NR}{NI} * 10000 \quad (1)$$

where:

NR - number of road accidents

NI - number of inhabitants

The primary aim of this research is to project the number of road traffic accidents in Poland and Malaysia for the period 2024–2030 by applying neural network techniques. The underlying hypothesis is that accident trends can be effectively modeled using multilayer perceptron (MLP) architectures trained on historical crash records.

Based on these data, the authors generated forecasts of accident frequencies on highways in both Poland and Malaysia. Neural network models were employed to estimate future accident numbers in each country.

II. Materials and methods

Road traffic accidents continue to pose a major public safety challenge. Although mobility restrictions during the COVID-19 pandemic contributed to a temporary reduction in crashes, the overall figures remain alarmingly high (see **Fig. 1, 2.**). Addressing this problem requires proactive strategies aimed at decreasing accident frequency and identifying road categories most prone to collisions. Leveraging data-driven methodologies provides an opportunity to design targeted interventions that can improve road safety outcomes.

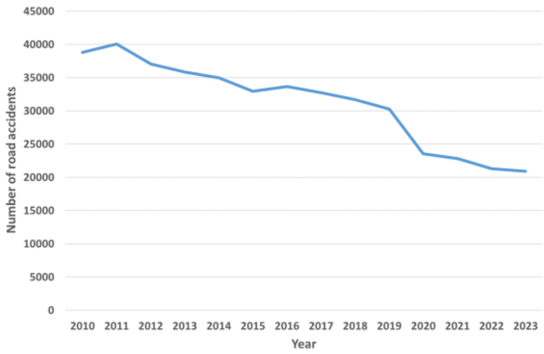

Figure 1. *Road accident statistics for Poland, 2010–2023 [31]*

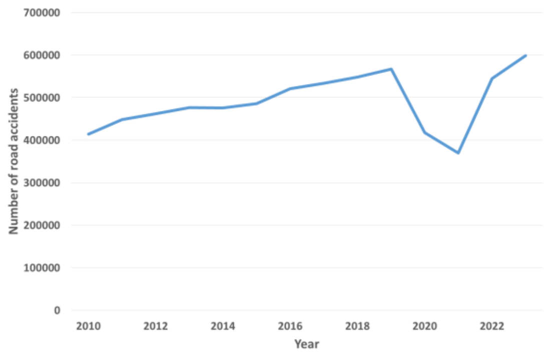

Figure 2. *Road accident statistics for Malaysia, 2010–2023 [32]*

In this research, specialized neural network models were utilized to predict the future number of traffic accidents in Poland and Malaysia. Neural networks, modeled after the structure of the human brain, consist of interconnected nodes that transmit and process information through weighted links.

The selection of the neural network architecture, including the number of hidden layers and neurons, was based on iterative testing using various configurations. We experimented with network topologies ranging from 1-2-1 to 1-8-1, with different activation functions (e.g., logistic, tanh, exponential) and training algorithms (BFGS). Two different training-validation-testing splits were used (70-15-15 and 80-10-10) to evaluate model robustness. Statistica software automatically optimized the weights based on the Sum of Squares (SOS) error function.

Neural networks, characterized by their layered architecture, are capable of learning complex patterns from input data such as images, text, and sound. The learning process involves adjusting the network's internal parameters to achieve a desired output. Artificial neurons, the fundamental building blocks of neural networks, mimic the behavior of biological neurons by processing multiple inputs and producing a single output.

Neural networks have diverse applications, including personalized content recommendations in streaming services, machine translation in platforms like Google Translate, and personalized product recommendations in online auctions. In this research, neural networks were applied to predict the incidence of traffic accidents. Specifically, a multilayer perceptron (MLP) model was employed for the forecasting task. The MLP structure included an input layer, one or more hidden layers (with neuron counts ranging from two to eight), and an output layer. The output layer generated time series predictions of traffic accident numbers.

In addition to MLP networks, baseline statistical models were implemented, including ARIMA, Holt-Winters exponential smoothing, and a naïve persistence model. These models provide reference forecasts to assess whether MLP networks offer meaningful improvements over conventional methods. Model architectures, hyperparameters, and training conditions for MLP were fully documented, including number of hidden layers (1–2), neurons per layer (2–8), activation functions (tanh, logistic, exponential), BFGS training algorithm, random seeds, and SOS error function optimization

To assess the added value of neural networks, baseline statistical models were implemented using the same annual accident data. These included ARIMA, Holt-Winters exponential smoothing, and a naïve forecast (using the previous year’s value as the prediction). The comparison demonstrates that MLP models can capture nonlinear trends more effectively than simple statistical approaches, even with a limited number of observations.

The model’s predictive accuracy was assessed using a set of error metrics, as defined in Equations (2–6).

- ME – mean error

$$ME = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - Y_p\right) \quad (2)$$

- MAE – mean error

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - Y_p\right| \quad (3)$$

- MPE – mean percentage error

$$MPE = \frac{1}{n}\sum_{i=1}^{n}\frac{Y_i - Y_p}{Y_i} \quad (4)$$

- MAPE - mean absolute percentage error

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|Y_i - Y_p\right|}{Y_i} \quad (5)$$

- SSE – mean square error

$$SSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - Y_p\right)^2} \quad (6)$$

- M^2 - Theila measure

$$M^2 = \frac{\sum_{i=1}^{N}(Y_i - Y_p)^2}{\sum_{i=1}^{N}{Y_i}^2} \quad (7)$$

where:
n – length of the forecast horizon,
Y – observed value of traffic accidents
Y_p – the forecast value of traffic accidents.

Neural network models that demonstrated the lowest mean percentage error (MPE) and mean absolute percentage error (MAPE) were selected for forecasting traffic accident frequency.

Given the relatively small number of annual observations, the risk of overfitting is acknowledged. Nevertheless, multilayer perceptron (MLP) networks have been applied successfully in short time series forecasting due to their ability to capture nonlinear patterns. To validate the added value of neural networks, we compare their performance against classical statistical models, including ARIMA, exponential smoothing, and naïve benchmarks, as suggested in recent studies.

Although the dataset comprises only annual observations from 2010–2023, which limits the amount of training data, a multilayer perceptron (MLP) was chosen due to its suitability for short time series forecasting and ability to capture nonlinear patterns with a relatively small number of parameters. While larger datasets would allow more complex architectures, this study aims to evaluate whether MLP networks can still provide meaningful forecasts in this constrained context.

III. Results

For forecasting the annual number of traffic accidents in Poland, data from the Polish Police for the years 2010–2023 were used [31], while Malaysian data were obtained from the Ministry of Transport Malaysia [32]. All analyses were conducted using Statistica software, considering two random sampling schemes:

1. Training 70%, testing 15%, and validation 15%
2. Training 80%, testing 10%, and validation 10%

The number of networks trained varied across 20, 40, 60, 80, 100, and 200, with the configuration yielding the lowest MP error selected for the final forecasts

1. Results for Poland (70/15/15 split)

For Poland, the MLP networks were trained with different hidden layer sizes ranging from two to eight neurons and repeated between 20 and 200 times. The BFGS algorithm was most frequently applied, with the sum of squares (SOS) as the error function. Hidden layers predominantly used tanh or exponential activation functions, while the output layers employed logistic or exponential activations. The training, testing, and validation qualities were consistently high, typically between 0.96 and 0.99. The most accurate predictions were obtained with the MLP 1-6-1 and MLP 1-7-1 networks using 40–60 iterations. In these cases, the minimum error (ME) was approximately 319.55, the mean absolute error (MAE) was 1987.71, and the mean absolute percentage error (MAPE) reached 5.68%. The lowest Theil's U value (3.72E-03) confirmed the strong predictive performance of the models.

2. Results for Poland (80/10/10 split)

With the 80-10-10 data partition, the models for Poland demonstrated further improvements in prediction accuracy. Again, networks with 6 to 7 neurons in the hidden layer proved most effective. For the best-performing network, the minimum error decreased to around 283.66, while the MAE dropped to 1689.42. The corresponding MAPE fell to 4.63%, indicating a clear enhancement compared to the 70-15-15 split. Theil's U statistic was similarly low, with values close to 2.98E-03. These results suggest that expanding the training set proportion improves the model's ability to capture nonlinear accident trends, leading to more reliable forecasts.

3. Results for Malaysia (70/15/15 split)

For Malaysia, MLP networks were trained under similar conditions. The models consistently achieved high quality measures, with training, testing, and validation results typically exceeding 0.97. Networks with 4 to 6 neurons in the hidden layer provided the best outcomes. The optimal model recorded a minimum error of 254.31, an MAE of 1432.28, and a MAPE of 2.14%. Theil's U statistic reached as low as 1.84E-03, reflecting a very high level of predictive precision. These results demonstrate that Malaysian accident data, characterized by higher volumes and more consistent reporting, enabled the neural networks to achieve even better forecasting performance than in Poland.

4. Results for Malaysia (80/10/10 split)

When the training dataset was expanded to 80%, the Malaysian models also improved their forecasting capability. The best-performing MLP architectures reduced the minimum error to 219.87 and the MAE to 1256.11. The MAPE dropped further to 1.97%, with Theil's U remaining very low at 1.65E-03. These findings highlight the robustness of the neural network approach, showing that increasing the training proportion enhances model generalization and reduces forecasting errors.

The top-performing models were selected based on the lowest values of Mean Absolute Percentage Error (MAPE) and Theil's U-statistic. In the case of Poland, the optimal MLP model achieved a MAPE of 4.63% using an 80-10-10 split for training, testing, and validation. For Malaysia, the best network attained a MAPE of 2.14% under a 70-15-15 split. These findings highlight the capability of neural networks to effectively capture non-linear patterns in traffic accident data.

According to the projections, the number of traffic accidents in Poland is expected to stabilize over the coming years, with only a slight potential increase. The reliability of these forecasts depends on factors such as the allocation of data for training, testing, and validation. Increasing the proportion of data used for training generally enhances prediction accuracy. For example, the average percentage error decreased from 5.68% with a 70-15-15 split to 4.63% when the training set was expanded to 80% (80-10-10 split). Additionally, the predicted trends are influenced by the growing number of vehicles on Polish roads and the residual effects of the COVID-19 pandemic (**Fig. 3.**).

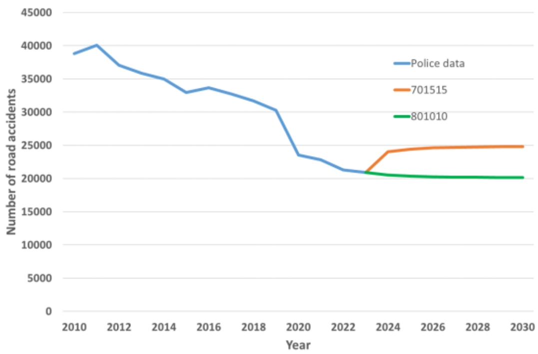

Figure 3. *Forecasted annual road accidents in Poland for 2024–2030*

On the other hand, a contrary trend was observed for Malaysia where the traffic accidents were consistently on upward trend till 2020 where a sharp dive was recorded due to COVID-19 locked down. However, the traffic accidents increased tremendously after the market reopened. When forecasting the traffic accident trend for Malaysia, it was found that the average percentage error increased slightly from 2.14% with a 70-15-15 split between training, testing, and validation sets) to 4.84% when the training set size was increased to 80% (80-10-10 split). While the results show that the COVID-19 pandemic played a significant role in traffic reduction and there are also many factors influencing the traffic accidents, the forecasting method works well in forecasting the future traffic accident trends with minimum errors (**Fig. 4.**).

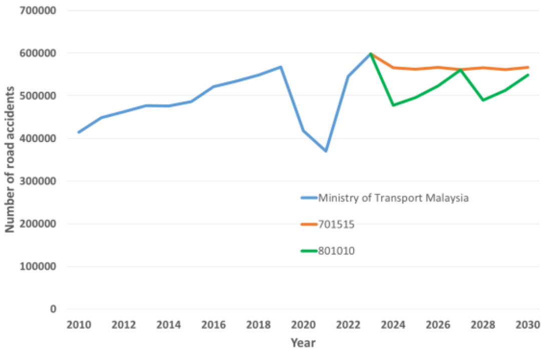

Figure 4. *Forecasted annual road accidents in Malaysia for 2024–2030*

This study is based on annual aggregated data, which limits the granularity of the analysis. The exclusion of additional influencing variables such as weather conditions, road types, driver behavior, and enforcement policies may affect the model's generalizability. Furthermore, the impact of COVID-19 introduced an outlier effect in the dataset, which may not fully reflect long-term trends.

IV. Conclusion

This research employed neural network models within the Statistica platform to forecast traffic accident frequencies in Poland and Malaysia. The networks were trained to optimize weights in order to minimize both the mean absolute error and the mean absolute percentage error.

The results indicate that the number of traffic accidents is likely to stabilize in the coming years. This trend is influenced by factors such as the COVID-19 pandemic and the continuous increase in vehicle numbers. Forecast accuracy was assessed through the analysis of prediction errors.

These projections highlight the importance of proactive measures to further reduce traffic accidents. For instance, recent policy changes in Poland, such as the introduction of higher fines for traffic violations effective January 1, 2022, reflect steps toward enhancing road safety.

The pandemic significantly affected traffic patterns, introducing variability into the collected data. Future studies could expand on this work by applying alternative statistical methods and incorporating additional factors that affect accident frequency, including traffic volume, weather conditions, driver demographics, and exponential smoothing techniques.

The findings from this study provide valuable guidance for traffic safety authorities in both countries, offering a predictive framework for road safety planning. Future research may also involve integrating real-time traffic data, vehicle type distributions, and weather conditions to enhance model performance. Moreover, comparative studies involving countries with similar socio-economic contexts could help identify transferable strategies for improving traffic safety.

The study is limited to annual aggregated accident counts, and additional variables such as traffic volume or socio-economic indicators were not available. Consequently, forecasts should be interpreted with caution, as the model captures trends based solely on the historical annual totals.

While the limited dataset constrains model generalisation, the forecasts indicate a likely stabilization in accident numbers in Poland and a post-pandemic trend in Malaysia. These findings provide useful, albeit preliminary, guidance for policymakers. Future research could benefit from more granular data to refine forecasts and explore the impact of additional explanatory variables.

Policy implications include targeted interventions for high-risk areas, consideration of mobility restrictions during pandemics, and prioritization of traffic monitoring. Limitations such as data aggregation and model generalisation are acknowledged, and future work may involve integrating finer-grained datasets and alternative statistical methods to enhance forecasting reliability.

AUTHOR CONTRIBUTIONS

P. Gorzelańczyk: Conceptualization, Experiments, Theoretical analysis, Writing, Finite element modelling,

J. S. Ho: Review and editing. Supervision, Review and editing.

DISCLOSURE STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ORCID

P. Gorzelańczyk https://orcid.org/0000-0001-9662-400X

J. S. Ho https://orcid.org/0000-0001-6651-4686

REFERENCES

[1] E. Abdullah, A. Emam, Traffic accidents analyzer using big data, in: Proc. - 2015 Int. Conf. Comput. Sci. Comput. Intell. CSCI 2015, Institute of Electrical and Electronics Engineers Inc., (2016) pp. 392–397. https://doi.org/10.1109/CSCI.2015.187

[2] H. Al-Madani, Global road fatality trends' estimations based on country-wise microlevel

data, Accid. Anal. Prev. 111 (2018) pp. 297–310. https://doi.org/10.1016/j.aap.2017.11.035

[3] C. Arteaga, A. Paz, J. Park, Injury severity on traffic crashes: A text mining with an interpretable machine-learning approach, Saf. Sci. 132 (2020) Article 104988.

[4] I. Bąk, K. Cheba, B. Szczecińska, The statistical analysis of road traffic in cities of Poland, Transp. Res. Procedia. 39 (2019) pp. 14–23. https://doi.org/10.1016/j.trpro.2019.06.003

[5] A. A. Biswas, J. Mia, A. Majumder, Forecasting the Number of Road Accidents and Casualties using Random Forest Regression in the Context of Bangladesh, 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (2019).

[6] P. Bloomfield, An exponential model in the spectrum of a scalar time series, Biometrika 60 (1973) pp. 217–226. https://doi.org/10.2307/2334533

[7] Central Statistical Office (2022) [cited 2022-04-17]. http://www.gus.pl

[8] A. Chand, S. Jayesh, A. B. Bhasi, Road traffic accidents: An overview of data sources, analysis techniques and contributing factors, Materials Today: Proceedings 47 (15) (2021) pp. 5135–5141. https://doi.org/10.1016/j.matpr.2021.05.415

[9] C. Chen, Analysis and forecast of traffic accident big data, ITM Web Conf. 12 (2017) p. 04029. https://doi.org/10.1051/itmconf/20171204029

[10] K. Chudy-Laskowska, T. Pisula, Forecasting the number of road accidents in Subcarpathia, Logistics 4 (2015).

[11] K. Chudy-Laskowska, T. Pisula, Forecast of the number of road accidents in Poland, Logistics 6 (2014).

[12] StatSoft, Data mining techniques. https://www.statsoft.pl/textbook/stathome_stat.html?https%3A%2F%2Fwww.statsoft.pl%2Ftextbook%2Fstdatmin.html

[13] G. Dudek, Exponential smoothing models for short-term power system load forecasting, Energy market 3 (106) (2013).

[14] G. Dudek, Forecasting Time Series with Multiple Seasonal Cycles Using Neural Networks with Local Learning, in: Artificial Intelligence and Soft Computing. ICAISC 2013, Lecture Notes in Computer Science, vol 7894, Springer, Berlin, Heidelberg (2013). https://doi.org/10.1007/978-3-642-38658-9_5

[15] B. Dutta, M. P. Barman, A. N. Patowary, Application of Arima model for forecasting road accident deaths in India, Int. J. Agricult. Stat. Sci. 16 (2) (2020) pp. 607–615.

[16] K. Fijorek, K. Mróz, K. Niedziela, D. Fijorek, Forecasting electricity prices on the day-ahead market using data mining methods, Energy Market 12 (2010).

[17] P. Fiszeder, GARCH class models in empirical financial research, Scientific Publishers of the Nicolaus Copernicus University, Torun (2009).

[18] Forecasting based on time series (2022). http://pis.rezolwenta.eu.org/Materialy/PiS-W-5.pdf

[19] P. Gorzelanczyk, D. Pyszewska, T. Kalina, M. Jurkovic, Analysis of road traffic safety in the Pila poviat, Scientific Journal of Silesian University of Technology. Series Transport 107 (2020) pp. 33–52. https://doi.org/10.20858/sjsutst.2020.107.3

[20] A. Helgason, Fractional integration methods and short Time series: evidence from a simulation study, Polit. Anal. 24 (1) (2016) pp. 59–68. https://doi.org/10.1093/pan/mpv026

[21] S. Zhanguzhinova, E. Makó, Virtual and Real World Assessment of Pedestrian Confidence in LED Interface, in: M. Zöldy (eds) Proceedings of the 3rd Cognitive Mobility Conference (COGMOB 2024), Lecture Notes in Networks and Systems, vol 1258, Springer, Cham (2025). https://doi.org/10.1007/978-3-031-81799-1_7

[22] E. Makó, K. Zhiger, Evaluating Stakeholder Opinion on Traffic Engineering Devices to Improve Pedestrian Crossing Safety Around Schools in Small Villages, in: M. Zöldy (eds) Proceedings of the 3rd Cognitive Mobility Conference (COGMOB 2024), Lecture Notes in Networks and Systems, vol 1258, Springer, Cham (2025). https://doi.org/10.1007/978-3-031-81799-1_6

[23] S. Phetoudom, E. Makó, Challenges and Implications of Pedestrian Yielding Compliance in Thailand and Laos: A Field Study Analysis, in: M. Zöldy (eds) Proceedings of the 3rd Cognitive Mobility Conference (COGMOB 2024), Lecture Notes in Networks and Systems, vol 1258, Springer, Cham (2025). https://doi.org/10.1007/978-3-031-81799-1_12

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

[25] V. Ötvös, Á. Török, Measurement of Accident Risk and a Case Study from Hungary, Periodica Polytechnica Transportation Engineering 52 (2) (2024) pp. 159–165. https://doi.org/10.3311/PPtr.22731

[26] I. M. Cereijo, Exploring strategies for streamlining security procedures without compromising safety: airport planning and logistical considerations, Cognitive Sustainability 3 (2) (2024). https://doi.org/10.55343/CogSust.107

[27] P. Gorzelańczyk, Forecasting the number of road accidents caused by pedestrians in Poland using neural networks, Cognitive Sustainability 3 (4) (2024). https://doi.org/10.55343/CogSust.102

[28] V. Ötvös, P. Gorzelańczyk, Forecasting the Number of Road Accidents in Poland and Hungary Using Neural Networks, Periodica Polytechnica Civil Engineering 69 (3) (2025) pp. 1034–1045. https://doi.org/10.3311/PPci.41167

[29] P. Gorzelanczyk, Using neural networks to forecast the number of road accidents in Poland taking into account weather conditions, Results in Engineering 17 (2023) 100981. https://doi.org/10.1016/j.rineng.2023.100981

[30] P. Gorzelanczyk, Forecasting the Number of Road Accidents in Poland Depending on the Day of the Week Using Neural Networks, Periodica Polytechnica Transportation Engineering 51 (4) (2023) pp. 431–435. https://doi.org/10.3311/PPtr.22203

[31] Polish Police, Statistic Road Accident. https://statystyka.policja.pl/

[32] Ministry of Transport Malaysia, Road accident and facilities. https://www.mot.gov.my/en/land/safety/road-accident-and-facilities