THE BALTIC JOURNAL
OF ROAD AND BRIDGE ENGINEERING
2026/21(2)

ISSN 1822-427X/eISSN 1822-4288
2026 Volume 21 Issue 2: 89–106
https://doi.org/10.7250/bjrbe.2026-21.679

USING AI MODEL TO ANTICIPATE ROAD ACCIDENTS OF POLAND AND KOSOVO

GËZIM HOXHA[1], PIOTR GORZELAŃCZYK[2,*]

[1]*Department of Traffic and Transport, Faculty of Mechanical Engineering, University of Pristina, Pristina, Kosovo*

[2]*Department of Transport, Stanislaw Staszic State University of Applied Sciences in Pila, Podchorazych 10 Street, 64-920 Pila, Poland*

Received 27 September 2025; accepted 22 May 2026

Abstract. This paper investigates the application of artificial neural networks for predicting road traffic accident counts in Poland and Kosovo. The study uses annual accident data collected for the period 2010–2023 from official statistical records provided by the Polish Police and the Kosovo Police. Forecasting procedures were carried out with multilayer perceptron (MLP) models developed in the Statistica environment. To assess the robustness of the models, two alternative dataset division schemes were employed, namely, 70%–15%–15% and 80%–10%–10% for the training, testing, and validation subsets. Forecast accuracy was measured using Mean Absolute Error (MAE) together with Mean Absolute Percentage Error (MAPE). The obtained forecasts suggest that the number of road traffic accidents in both analysed countries is likely to remain relatively stable over the 2024–2030 period. The analysis further demonstrates that a larger share of training observations contributes to improved predictive performance and lower estimation errors. Nevertheless, the relatively short time series constitutes a methodological limitation and requires careful interpretation of the forecasting results.

Keywords: forecasting, Kosovo, neural networks, pandemic, Poland, traffic accident.

* Corresponding author. E-mail: piotr.gorzelanczyk@ans.pila.pl

Gëzim HOXHA (ORCID ID 0000-0002-0486-7010)
Piotr GORZELAŃCZYK (ORCID ID 0000-0001-9662-400X)

Introduction

Road traffic accidents remain one of the most serious global public safety challenges, resulting not only in substantial material losses but also in severe injuries and fatalities among drivers, passengers, and pedestrians. According to the World Health Organization, approximately 1.3 million people lose their lives in road crashes every year, making traffic accidents one of the leading causes of mortality worldwide (Abdullah et al., 2015; Al-Madani, 2018). Their economic consequences are equally significant, as the average national cost associated with road crashes is estimated at nearly 3% of gross domestic product. Among children and young adults aged 5–29 years, road accidents constitute the primary cause of death. In response to this challenge, the United Nations General Assembly established a target to reduce the number of deaths and injuries caused by traffic accidents by 50% by 2030.

The impact of a road collision is largely determined by its severity, which directly influences the scale of social and economic losses. Reliable estimation of accident severity is therefore essential for public authorities responsible for transport policy and road safety management. Accurate assessments support the development of preventive regulations and safety interventions designed to reduce fatalities, injuries, and infrastructure damage (Al-Madani, 2018; Arteaga et al., 2020). Effective implementation of such measures requires identification of the most influential determinants of accident severity (Bąk et al., 2019). Recent advances in computational modelling have enabled the application of deep learning methods, such as multi-node Deep Neural Networks, which can classify and forecast different severity levels of traffic incidents, including injury occurrence, fatal outcomes, and financial losses (Biswas et al., 2019).

Road accident statistics are typically derived from multiple institutional sources. Public agencies usually collect and process these data through police documentation systems, insurance company records, and healthcare databases. This integrated data collection framework enables a comprehensive examination of accident patterns and improves the analytical capacity of transport safety studies (Bloomfield, 1973).

Modern intelligent transportation systems have become an increasingly important source of traffic-related information for both accident analysis and forecasting. Data collected from in-vehicle GPS systems and automated traffic monitoring technologies allow for continuous observation of traffic conditions and support predictive modelling approaches (Statistics Poland, 2022; Hoxha et al., 2023a). Additional information can be obtained through roadside microwave detection devices, which record vehicle classifications, traffic density, and travel speeds in real time (Chand et al., 2021; Hoxha et al., 2023b). Similarly, automated license plate recognition systems facilitate large-scale traffic flow monitoring over specified periods (Chudy-Laskowska & Pisula, 2015). Social media platforms may also provide supplementary information regarding accidents and congestion, although the reliability of such reports is often limited by the non-professional nature of the reporting process (Chudy-Laskowska & Pisula, 2014).

The credibility of traffic accident analysis depends strongly on rigorous validation procedures and cross-verification of heterogeneous data sources. Integrating datasets originating from multiple systems significantly improves analytical consistency and increases the robustness of statistical conclusions (Chudy-Laskowska & Pisula, 2014).

Previous studies have examined road safety from different analytical perspectives. For example, Vilaca et al., 2017 conducted statistical investigations into relationships between traffic participants and accident severity, concluding that stronger regulatory enforcement and improved safety policies are required. Similarly, Bąk et al. (2019) analysed road safety conditions in a selected Polish region using accident frequency as a key explanatory indicator, while applying multivariate statistical methods to identify behavioural and demographic factors associated with accident responsibility (Dudek et al., 2013a). The selection of an appropriate accident database depends on the nature of the traffic safety problem being investigated. Combining traditional statistical models with real-world driving data and information collected from intelligent transport systems substantially improves predictive precision and supports more effective accident prevention strategies (Dudek et al., 2013b).

Various methodological approaches have been proposed for forecasting traffic accident occurrence. Time series models remain among the most frequently applied techniques for estimating accident frequency (Dutta et al., 2020; Fijorek et al., 2010). However, these approaches often suffer from residual autocorrelation problems (Fiszeder, 2009) and may not always allow for effective validation using prior forecast performance. Alternative forecasting methods include exponential smoothing techniques, such as the Holt-Winters model, which has been successfully applied in road accident prediction, including extensions involving multi-seasonality structures (Gorzelańczyk et al., 2024). Nevertheless, these methods are limited by their inability to incorporate exogenous explanatory variables directly into the forecasting process (Helgason, 2016).

According to data published by the Western Balkans Road Safety Observatory, Kosovo recorded the highest number of fatal road accident victims in the 45–59 age category in 2023. Despite this fact, the total number of fatalities across all age groups remains moderate compared with other Western Balkan countries, suggesting relative stability in the regional road safety context (WBRSO, 2023).

Based on data on the treated road accident cases by Kosovo Courts, the number of accidents continues to be at a high level (Hoxha et al., 2023b).

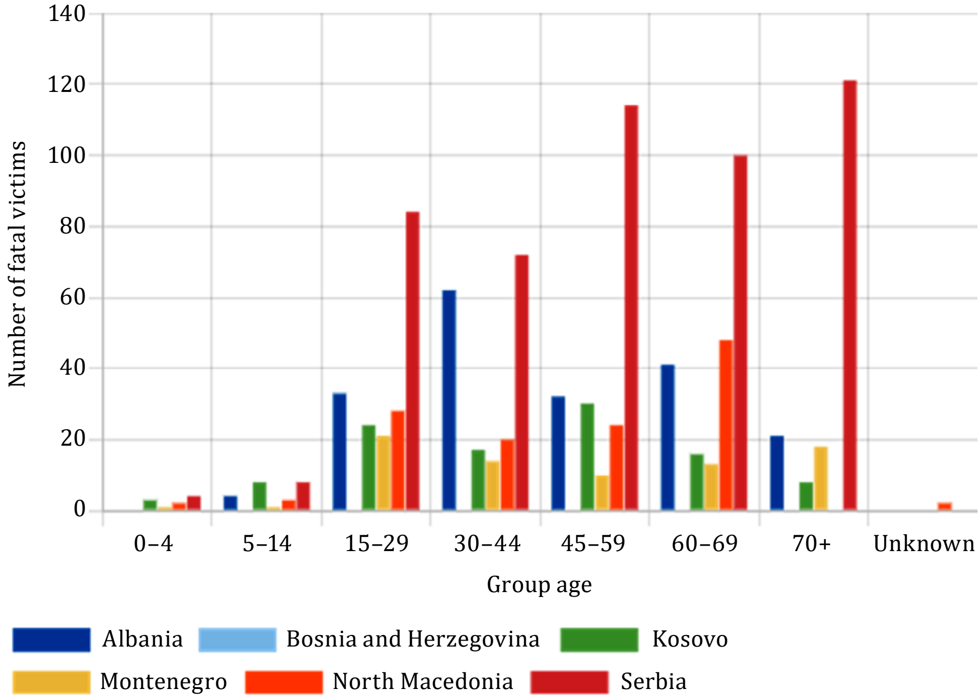

Figure 1. Number of fatal victims by group age for Kosovo in 2023
(Source: Western Balkans Road Safety Observatory)

The number of accidents per 10 000 population (*NRA*) is another crucial factor to consider (Gorzelańczyk et al., 2025). There were 20 936 traffic incidents in Poland in 2023, a country with 37.6 million people. In 2023, there were 5.57 traffic accidents per 10 000 people in Poland, according to the data included into Equation (1). During the same time period, Kosovo had 1.92 million residents and 23 676 road accidents. In this instance, Kosovo had 22 times more traffic accidents per 10 000 inhabitants in 2023 (123.3) than Poland.

$$NRA = \frac{NR}{NI} \times 10000, \tag{1}$$

where NR – the number of traffic accidents;
NI – the number of residents.

The authors estimated the number of accidents on Polish and Kosovar highways based on the aforementioned data. The number of accidents in both nations was predicted using neural networks.

2. Methodological approach

There are a lot of traffic accidents on the roadways each year. Road accidents have decreased in recent years due to the pandemic, which has an impact on the forecast value received. Even with the pandemic, there are still a lot of traffic accidents. Because of this, every effort must be made to lower this figure and determine which kinds of roads will see the most traffic accidents (Figure 2).

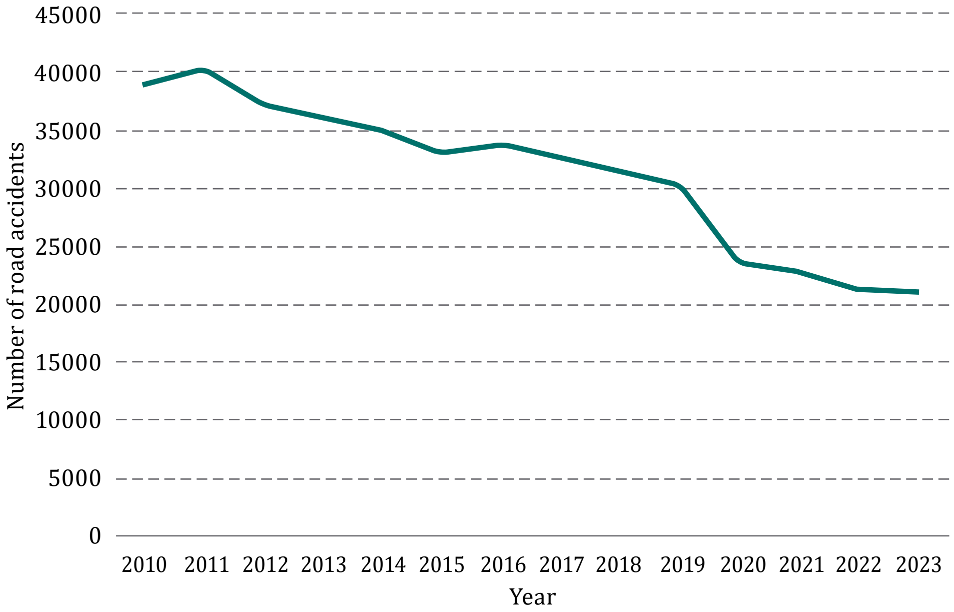

Figure 2. Poland's number of traffic accidents from 2010 to 2023
(Source: Statistic Road Accident https://statystyka.policja.pl/)

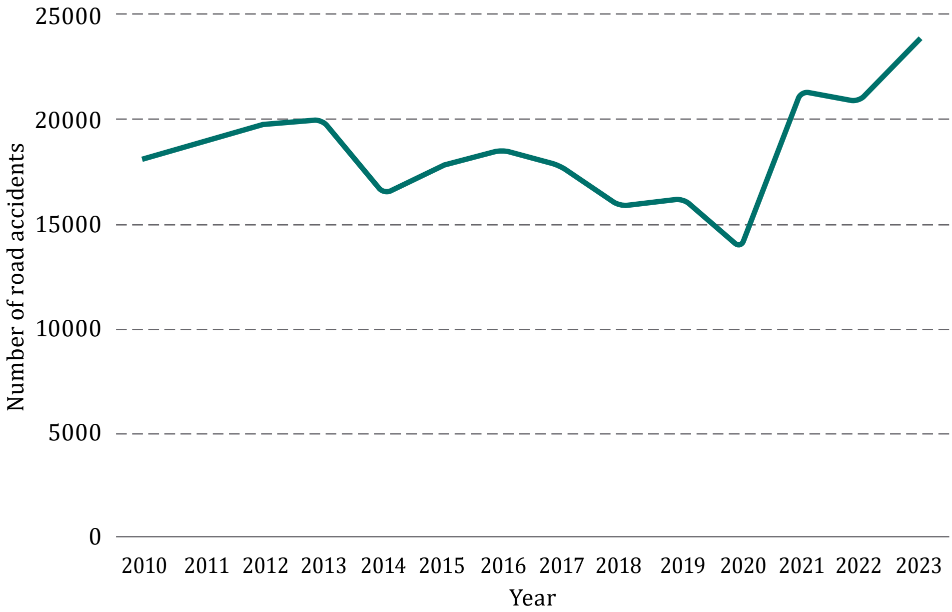

Figure 3. Kosovo's traffic accident rate from 2010 to 2023
(Source: Kosovo Police https://www.kosovopolice.com)

In this study, artificial neural networks were applied to forecast the number of road traffic accidents in Poland and Kosovo. The adopted approach was based on computational models designed to imitate selected mechanisms of human brain operation. A neural network is composed of interconnected processing units (neurons) linked through weighted connections, where each neuron receives input data, processes it, and generates an output signal. The forecasting capability of the network depends largely on the structure of the model and the optimisation of its parameters.

Artificial neural networks can be formally defined as mathematical systems inspired by biological nervous systems. Their architecture generally consists of interconnected layers responsible for data transmission and transformation. During the training process, the network receives input information such as numerical values, text, images, or sound signals and adjusts internal parameters to minimise prediction error. Artificial neurons constitute the fundamental computational elements of such systems and perform functions analogous to biological neurons by transforming multiple input signals into a single output value (Hoxha et al., 2023a; Gorzelańczyk et al., 2024). Neural networks are considered one of the core technologies underlying modern artificial intelligence because they enable pattern recognition, learning, and generalization based on historical observations (Lake et al., 2017).

Neural network methods are currently used across numerous application domains. Examples include recommendation systems used by streaming platforms, personalized product selection in e-commerce systems, automated language translation services such as Google Translate, and pattern recognition tasks in intelligent information systems (Becoming Human, 2019; Forbes, 2020; Wu et al., 2020; Oronowicz-Jaśkowiak, 2019). Forecasting problems, including the prediction of road accident frequency, also constitute an important area of application for neural network models.

The forecasting procedure implemented in this study employed multilayer perceptron (MLP) neural networks available in the Statistica software package. The optimisation process performed by the software involved selecting weight values that minimised forecasting error during the learning stage (Gorzelańczyk et al., 2024). The adopted MLP architecture contained one hidden layer with a variable number of neurons ranging from two to eight. The output layer consisted of a single neuron representing the predicted value of the accident time series. Different network structures were tested to identify configurations providing the best predictive performance.

Forecast quality was evaluated using prediction error measures calculated for the analysed models. As the investigated problem is based on annual time series characterised by a relatively small number of observations, several methodological limitations should be emphasised. In small datasets, artificial neural networks may be susceptible to overfitting, which can reduce model generalization ability and forecasting stability. For this reason, the neural network approach adopted in the study should be interpreted primarily as an exploratory analytical tool intended to assess the capability of nonlinear models to capture accident trends. In practical forecasting applications involving short time series, traditional statistical approaches such as ARIMA models or exponential smoothing methods are often considered more appropriate. Consequently, the results obtained from the neural network models should be treated as complementary rather than conclusive.

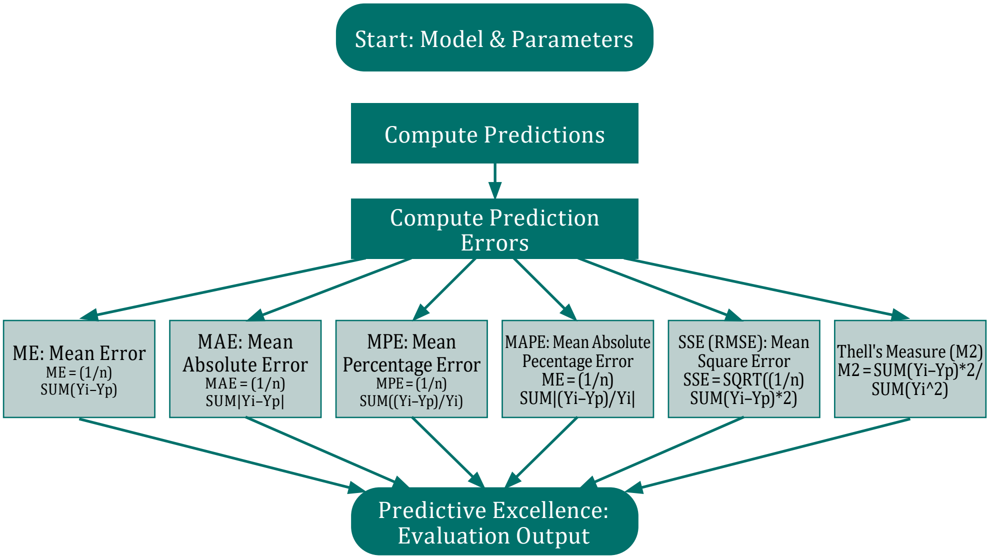

Figure 4. MLP neural network architecture and forecasting procedure

To predict the probability of traffic accidents in dependency, neural network models with the lowest average percentage error and average absolute percentage error were used.

3. Results

The forecasting analysis for Poland was performed using annual road accident statistics obtained from the Polish Police database for the years 2010–2023 (Statistic Road Accident Poland, 2024). In the case of Kosovo, the study utilised data published by the Kosovo Police (www.kosovopolice.com, 2024). For each country, the dataset contained 14 annual observations describing the recorded number of traffic accidents during the analysed period.

The research problem was formulated as a univariate time series forecasting task. In the adopted modelling framework, time expressed in years constituted the independent variable, whereas the dependent variable represented the annual number of road traffic accidents. The models were developed without the inclusion of additional explanatory or exogenous variables.

The computational analysis was conducted using the Statistica software environment (Gorzelańczyk et al., 2024). To evaluate model performance under different learning conditions, two alternative data partitioning schemes were applied:

1. 70% of observations for training, 15% for testing, and 15% for validation;
2. 80% of observations for training, 10% for testing, and 10% for validation.

An important methodological limitation of the study concerns the application of random sampling procedures to time series data. Random allocation of observations into training, testing, and validation subsets disrupts the chronological structure of the series and may introduce information leakage between datasets. Consequently, the obtained forecasting accuracy measures can be overly optimistic and should therefore be interpreted carefully. Future studies should employ validation approaches specifically designed for time-dependent data, including chronological splitting procedures or rolling-origin forecasting techniques.

For each analysed scenario, multiple neural network configurations were tested. The number of generated learning networks equal led 20, 40, 60, 80, 100, and 200, respectively, and the final model selection was based on the minimum MP error criterion (Statistic Road Accident Poland, 2024; Gorzelańczyk et al., 2024). Detailed modelling results are presented in Tables 1–2.

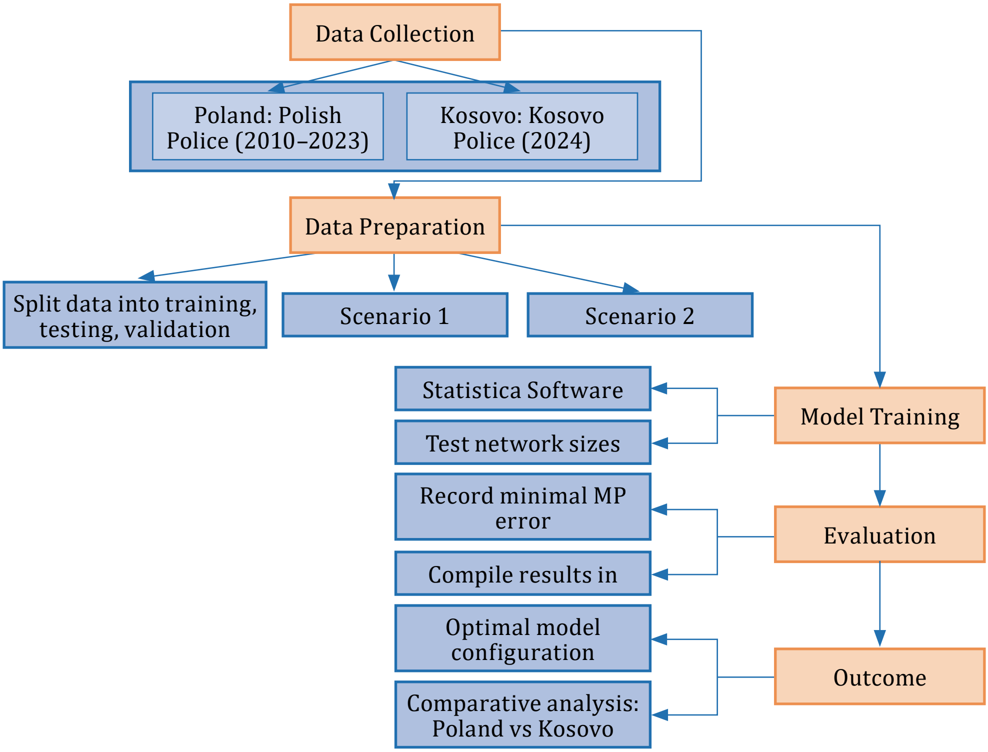

Figure 5. Methodological approach for traffic accident prediction using AI models

Table 1. Summary of neural network learning for the case of random sample size teaching 70%, testing 15% and validation 15% for Kosovo

Network number	Quality (learning)	Learning algorithm	Error function	Hidden layer activation	Output activation function	Errors					
						ME	MAE	MPE	MAPE	SSE	Theill
20	0.654247	B 13	S	T	L	380.81	1165.44	0.031436	6.79%	1684.57	3.45e-03
20	0.256850	B 0	S	L	Exp.	147.34	1515.86	0.023264	8.51%	2179.96	3.89e-03
20	0.164666	B 0	S	Exp.	Exp.	308.92	1710.53	0.000577	9.25%	2347.02	1.43e-01
20	0.460403	B 17	S	T	L	80.86	1239.53	0.016904	7.04%	1833.60	1.49e-01
20	0.677185	B 21	S	Exp.	T	114.81	1081.11	0.01394	6.17%	1619.26	4.39e-03
40	0.509204	B 20	S	Exp.	L	132.03	1179.97	0.018565	6.77%	1714.73	3.45e-03
40	0.140981	B 2	S	L	L	30.11	1607.62	0.017827	8.92%	2303.49	3.89e-03
40	0.564086	B 28	S	Exp.	T	148.95	1537.93	0.004711	8.45%	2121.83	1.43e-01
40	0.368610	B 21	S	T	Exp.	126.45	1328.20	0.020634	7.51%	2012.97	1.49e-01
40	0.354329	B 6	S	Exp.	L.	158.28	1312.42	0.021187	7.46%	1932.13	4.39e-03
60	0.524375	B 22	S	L	L	86.05	1303.14	0.017184	7.37%	1841.30	3.45e-03
60	0.509658	B 9999	S	Exp.	L	48.46	1220.47	0.01454	6.92%	1766.49	3.89e-03
60	0.718703	B 22	S	L	T	230.44	1385.36	0,001086	7.49%	2063.47	1.43e-01
60	0.502819	B 0	S	L	Exp.	91.38	1496.36	0.020342	8.39%	2161.24	1.49e-01
60	0.656495	B 23	S	Exp.	L	18.44	1051.81	0.00805	5.95%	1581.72	4.39e-03
80	0.688262	B 26	S	Exp.	L	54.81	1063.22	0.011925	6.04%	1590.83	3.45e-03
80	0.512565	B 21	S	Exp.	Exp.	119.52	1246.61	0.019457	7.09%	1860.94	3.89e-03
80	0.457815	B 28	S	T	L	161.62	1384.92	0.004548	7.63%	1978.99	1.43e-01
80	0.398456	B 0	S	Exp.	Exp.	5.88	1468.78	0.014475	8.15%	2130.28	1.49e-01
80	0.132890	B 3	S	L	L	61.76	1616.35	0.019629	8.99%	2307.52	4.39e-03
100	0.568222	B 20	S	T	Exp.	95.83	1296.91	0.017345	7.33%	1839.30	3.45e-03
100	0.675196	B 19	S	T	L	346.58	915.81	0.02827	5.46%	1421.02	3.89e-03
100	0.744127	B 3	S	L	T	48.54	1607.71	0.011549	8.83%	2179.71	1.43e-01
100	0.133666	B 3	S	L	L	95.82	1631.83	0.021656	9.10%	2320.19	1.49e-01
100	0.170616	B 0	S	L	Exp.	11.91	1568.99	0.016412	8.70%	2262.83	4.39e-03
200	0.602771	B 26	S	L	T	183.22	1226.71	0.000698	6.78%	1728.53	3.45e-03
200	0.712036	B 28	S	Exp.	L	200.78	1058.13	0.02038	6.10%	1609,54	3.89e-03
200	0.578844	B 23	S	T	L	133.82	1329.33	0.005001	7.36%	1959.46	1,43e-01
200	0.557312	B 31	S	L	L	212.73	1171.40	0.022514	6.76%	1701.02	1.49e-01
200	0.140981	B 3	S	L	L	105.08	1624.49	0.022069	9.07%	2308.88	4.39e-03
					Minimal	**5.88**	**915.81**	**0.06%**	**5.95%**	**1421.02**	**3.45e-03**

Table 2. Summary of neural network learning for the case of random sample size teaching 80%, testing 10% and validation 10% for Kosovo

Network number	Training quality	Learning algirthm	Function	Hidden layer activation	Output activation function	Errors					
						ME	MAE	MPE	MAPE	SSE	Theill
20	0.408585	B 4	S	Exp.	L	378.58	1483.31	0.034903	8.48%	2235.31	3.45e-03
20	0.571507	B 14	S	T	L	728.66	1481.24	0.052993	8.69%	2067.54	3.89e-03
20	0.320108	B 3	S	L	T	293.80	1752.60	0.032386	9.90%	2361.86	1.43e-01
20	0.447904	B 8	S	Exp.	L	133.04	1423.16	0.020456	8.00%	2139.61	1.49e-01
20	0.734638	B 40	S	Exp.	T	259.74	1048.70	0.022117	6.06%	1575.98	4.39e-03
40	0.736413	B 25	S	Exp.	T	239.54	1057.25	0.021325	6.11%	1601.36	3.45e-03
40	0.741959	B 46	S	L.	T	297.46	1055.51	0.024409	6.13%	1595.22	3.89e-03
40	0.738185	B 50	S	L.	L	51.47	1441.89	0.007896	7.85%	2102.66	1.43e-01
40	0.738699	B 42	S	T	T	236.19	1074.72	0.021161	6.22%	1623.51	1.49e-01
40	0.736267	B 50	S	T	Exp.	265.64	1047.03	0.022458	6.06%	1581.65	4.39e-03
60	0.735168	B 26	S	Exp.	L	285.09	1040.15	0.023508	6.03%	1577.52	3.45e-03
60	0.705624	B 20	S	T	L	158.61	1053.83	0.015944	6.04%	1564.26	3.89e-03
60	0.738085	B 22	S	T	L	14.28	1410.53	0.009399	7.72%	2027.88	1.43e-01
60	0.667801	B 57	S	Exp.	T	279.60	1172.53	0.025173	6.80%	1676.00	1.49e-01
60	0.721921	B 41	S	T	L	253.35	1079.94	0.022925	6.23%	1659.64	4.39e-03
80	0.413231	B 30	S	Exp.	Exp.	334.44	1478.36	0.032839	8.44%	2209.90	3.45e-03
80	0.442552	B 4	S	Exp.	L	456.12	1498.89	0.038562	8.58%	2212.82	3.89e-03
80	0.703489	B 29	S	T	T	190.08	1485.23	0.001475	8.10%	2112.06	1.43e-01
80	0.363277	B 6	S	L	L	217.78	1525.68	0.026937	8.60%	2249.83	1.49e-01
80	0.747711	B 37	S	Exp.	Exp.	256.20	1073.23	0.021569	6.21%	1569.90	4.39e-03
100	0.461736	B 16	S	L.	T	619.80	1570.31	0.048886	9.14%	2246.49	3.45e-03
100	0.383269	B 14	S	Exp.	Exp.	271.49	1486.62	0.029365	8.43%	2242.39	3.89e-03
100	0.349845	B 3	S	L	L	5.83	1740.65	0.015939	9.63%	2332.02	1.43e-01
100	0.331946	B 2	S	L	T	556.55	1761.96	0.047304	10.13%	2385.43	1.49e-01
100	0.415527	B 9	S	T	L	279.92	1462.82	0.029554	8.32%	2200.58	4.39e-03
200	0.661850	B 39	S	T	L	287.27	1175.18	0.025638	6.82%	1683.04	3.45e-03
200	0.518117	B 28	S	L.	T	382.74	1410.48	0.034016	8.08%	2050.14	3.89e-03
200	0.315648	B 3	S	L	T	165.18	1756.22	0.025133	9.82%	2353.99	1.43e-01
200	0.712530	B 19	S	Exp.	L	383.31	1123.13	0.029787	6.54%	1601.21	1.49e-01
200	0.678125	B 41	S	T	L	217.75	1055.79	0.021203	6.15%	1632.18	4.39e-03
					Minimal	**5.83**	**1040.15**	**0.15%**	**6.03%**	**1564.26**	**3.45e-03**

The forecasting results suggest that the number of road traffic accidents in Poland is expected to remain relatively stable over the coming years, although a moderate increase in accident frequency may still occur. The obtained forecasts indicate a tendency toward stabilization rather than substantial long-term growth in accident numbers.

The analysis also demonstrates that the adopted data partitioning strategy has a direct influence on forecasting accuracy. Comparison of the two examined variants, namely 70%–15%–15% and 80%–10%–10% for training, testing, and validation subsets, respectively, showed that enlarging the training dataset improved the predictive capability of the neural network models. In the case of Poland, the Mean Absolute Percentage Error (MAPE) decreased from approximately 5.68% under the 70–15–15 configuration to 4.63% when the 80–10–10 division was applied. Comparable behaviour was observed for the Kosovo dataset.

This phenomenon is particularly important in the context of limited time series length. With only a small number of observations available, each additional case included in the training subset significantly affects the learning process and the network's ability to approximate the underlying trend. At the same time, reducing the number of observations assigned to testing and validation subsets weakens the reliability of model evaluation and may lead to less robust estimates of predictive performance.

The results therefore indicate that lower forecasting errors can be achieved by increasing the proportion of observations used during the training stage. However, this improvement comes at the cost of reduced evaluation stability due to smaller testing and validation samples. In the analysed models, the forecasting error reached approximately 5.68% for the dataset partitioned according to the 70–15–15 structure, while the second configuration (80–10–10) reduced the error to approximately 4.63%.

The observed forecasting patterns may additionally be associated with external factors influencing traffic conditions in Poland during recent years. In particular, the systematic increase in the number of registered vehicles as well as disruptions related to the COVID-19 pandemic may have contributed to changes in traffic intensity and accident occurrence (Figure 6).

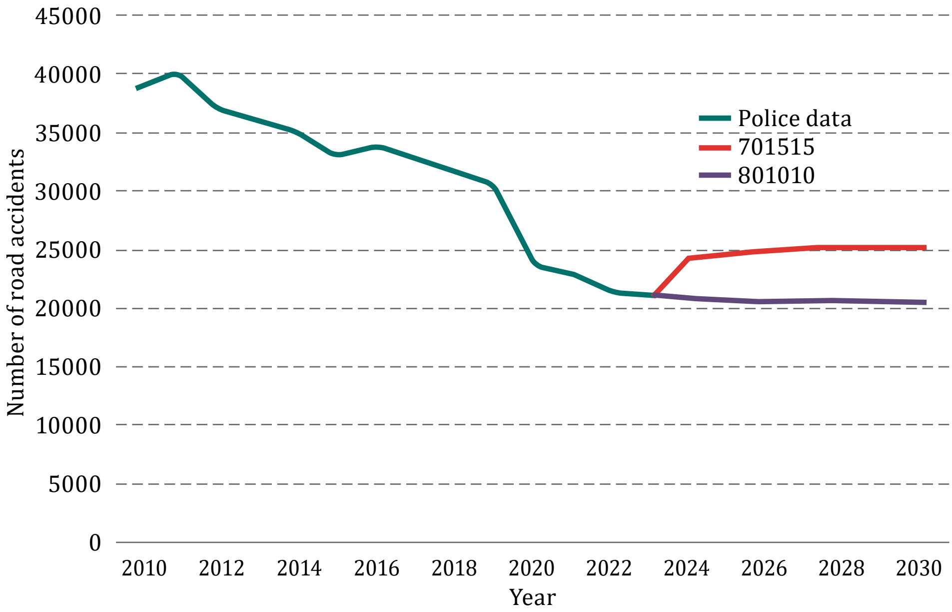

Figure 6. Estimated number of road accidents in Poland for the period 2024-2030

The obtained forecasting results indicate that the number of road traffic accidents in Kosovo is expected to remain relatively stable in the coming years, with a possible slight downward trend in accident frequency. The forecasts therefore suggest a gradual stabilization of the road safety situation rather than substantial fluctuations in accident occurrence.

As in the case of Poland, the adopted data partitioning strategy significantly influenced the predictive performance of the neural network models. The findings demonstrate that increasing the proportion of observations assigned to the training subset contributes to lower forecasting error values. However, because the analysed dataset contains only a limited number of annual observations, reducing the size of the testing and validation subsets decreases the robustness and reliability of model assessment.

The Kosovo results confirm the same methodological pattern identified for the Polish dataset. A larger training subset improves the network's ability to learn the underlying structure of the time series, but simultaneously weakens the statistical reliability of the validation process. Under the 70%–15%–15% partitioning scheme, the MAPE reached approximately 5.46%. Similar to the Polish analysis, the reduction of testing and validation observations contributed to improved fitting performance but increased the risk of overestimating model accuracy.

The obtained results may additionally reflect external factors affecting road traffic conditions in Kosovo during recent years. In particular, the continuous increase in the number of vehicles registered on Kosovo's roads, together with disruptions associated with the COVID-19 pandemic, may have influenced traffic intensity and accident dynamics over the analysed period (Figure 7).

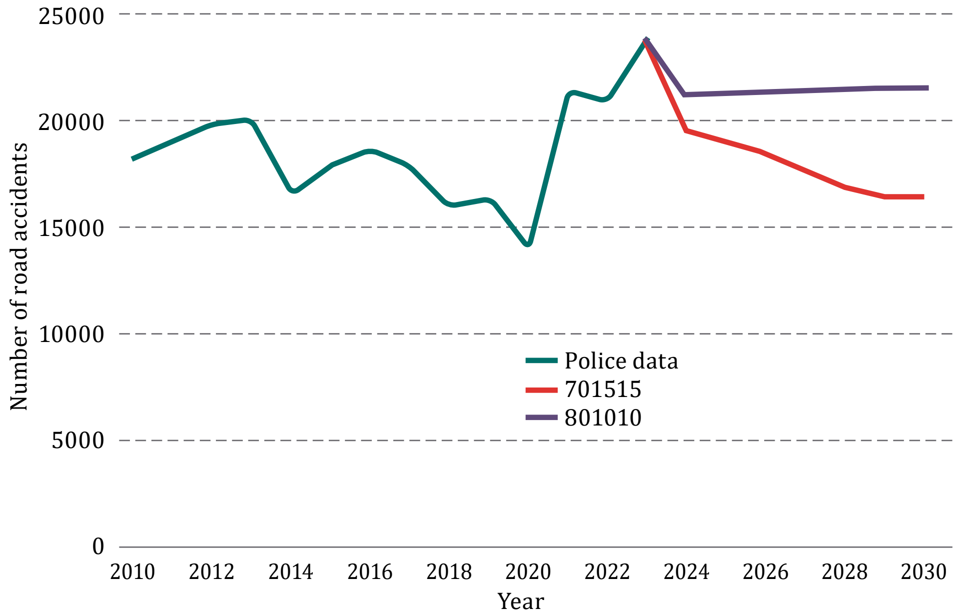

Figure 7. Estimated number of road accidents for the period 2024-2030 in Kosovo

Conclusion

The analysis was performed using the Statistica computer program where artificial neural networks were used in an effort to forecast the number of accidents on Polish and Kosovo roads. In the training phase of the neural network models, the algorithm would adjust the weights of the network with an aim of reducing forecast errors such as MAE and MAPE.

The analysis employed annual accident data and applied artificial neural networks as a forecasting tool for estimating future accident trends in both countries. The obtained results indicate a possible stabilization in the number of road accidents within the analysed forecast horizon. Nevertheless, the interpretation of these findings requires caution because of several important methodological constraints.

A key limitation of the study is the relatively small dataset, consisting of only 14 annual observations for each country. In addition, the use of random data partitioning for training, testing, and validation purposes in a time series framework may reduce the reliability of model evaluation and increase the risk of information leakage between subsets. Consequently, the forecasting accuracy measures reported in the study may be optimistic and should not be interpreted as conclusive evidence of predictive effectiveness.

For this reason, the generated forecasts should be considered exploratory and indicative rather than definitive. Future studies should focus on extending the temporal range of the datasets, incorporating additional explanatory variables related to traffic conditions and socioeconomic factors, and applying validation procedures specifically designed for time series analysis, such as chronological splitting or rolling forecasting origin methods. Such improvements would increase both the robustness and the generalizability of the forecasting models.

Declarations

Funding

This research received no external funding.

Author Contributions

Conceptualization, P.G.; methodology, P.G.; software, P.G.; validation P.G and G.H.; formal analysis, G.H.; investigation, P.G and G.H.; resources, P.G and G.H.; data curation, P.G and G.H.; writing – original draft preparation, P.G.; writing – review and editing, G.H.; visualisation, P.G. and G.H. All authors have read and agreed to the published version of the manuscript.

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.

Statement of the Use of Generative AI and AI-assisted Technologies in the Writing Process

Generative AI and AI-assisted technologies were used solely for language editing, grammar correction, and improvement of technical writing. All content was reviewed and verified by the authors, who take full responsibility for the final manuscript.

REFERENCES

Abdullah, E., & Emam A. (2015). Traffic accidents analyzer using big data. *2015 Int. Conf. Comput. Sci. Comput. Intell. CSCI 2015*, Las Vegas, USA, 392–397. https://doi.org/10.1109/CSCI.2015.187

Al-Madani, H. M. N. (2018). Global road fatality trends' estimations based on country-wise microlevel data. *Accident Analysis & Prevention*, *111*, 297–310. https://doi.org/10.1016/j.aap.2017.11.035

Arteaga, C., Paz, A. & Park, J. W. (2020). Injury severity on traffic crashes: A text mining with an interpretable machine-learning approach. *Safety Science*, *132*, Article 104988. https://doi.org/10.1016/j.ssci.2020.104988

Bąk, I., Cheba, K., & Szczecińska, B. (2019). The statistical analysis of road traffic in cities of Poland. *Transportation Research Procedia*, *39*, 14–23. https://doi.org/10.1016/j.trpro.2019.06.003

Becoming Human. (2019, August 10). *How Netflix uses AI, data science, and machine learning – from a product perspective*. www.becominghuman.ai

Biswas, A. A., Mia, J., & Majumder, A. (2019). 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)*, Kanpur, India, 1–5. https://doi.org/10.1109/ICCCNT45670.2019.8944500

Bloomfield, P. (1973). An exponential model in the spectrum of a scalar time series. *Biometrika*, *60*(2), 217–226. https://doi.org/10.1093/biomet/60.2.217

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

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

Chudy-Laskowska, K., & Pisula, T. (2015). Forecasting the number of road accidents in Subcarpathia. *Logistics*, *4*, 2782–2796.

Chudy-Laskowska, K., & Pisula, T. (2014). Forecast of the number of road accidents in Poland. *Logistics*, *6*. https://doi.org/10.14513/actatechjaur.00951

StatSoft. (1997). *Data mining techniques*. StatSoft https://www.statsoft.pl/textbook/stathome_stat.html?https%3A%2F%2Fwww.statsoft.pl%2Ftextbook%2Fstdatmin.html

Dudek, G. (2013a). Exponential smoothing models for short-term power system load forecasting. *Energy Market*, *106*(3), 14–19.

Dudek, G. (2013b). Forecasting time series with multiple seasonal cycles using neural networks with local learning. In: Rutkowski, L., Korytkowski, M., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (Eds.), *Artificial Intelligence and Soft Computing. ICAISC 2013. Lecture Notes in Computer Science*, vol. 7894. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-38658-9_5

Dutta, B., Barman, M. P., & Patowary, A. N. (2020). Application of Arima model for forecasting road accident deaths in India. *International Journal of Agricultural and Statistical Sciences*, *16*(2), 607–615. https://connectjournals.com/pages/articledetails/toc032612

Fijorek, K., Mróz, K., Niedziela, K., & Fijorek, D. (2010). Forecasting electricity prices on the day-ahead market using data mining methods. *Energy Market*, *91*(6) 46–50.

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

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

Forbes. (2020, September 12). *The amazing ways eBay is using Artificial Intelligence to boost business success*. www.forbes.com

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

Gorzelańczyk, P., & Ho, J. S. (2024). Forecasting the number of road accidents in Poland by road type. *Highlights of Vehicles*, *2*(1), 13–23. https://doi.org/10.54175/hveh2010002

Helgason, A. (2016). Fractional integration methods and short time series: evidence from a simulation study. *Political Analysis*, *24*(1), 59–68. https://doi.org/10.1093/pan/mpv026

Hoxha, G., Bixhaku, M., & Duraku, R. (2023a). Developing a new model for assessment of heavy vehicle-pedestrian collisions. *The Baltic Journal of Road and Bridge Engineering*, *18*(3), 102–123. https://doi.org/10.7250/bjrbe.2023-18.610

Hoxha, G., Fandaj, A., & Bajrami, X. (2023b). Quality of automatic traffic volume counting by cameras and impact on the qualitative indicators of traffic. *Infrastructures*, *8*(3), Article 44. https://doi.org/10.3390/infrastructures8030044

Hoxha, G., Gorzelanczyk, P., Likaj, R., Thaqi, S., & Cikaqi, F. (2025). Smart traffic management at intersections. *Journal of Sustainable Construction Materials and Technologies*, *10*(2), Article 5. https://doi.org/10.29187/2458-973X.1187

Kosovo Police. (2024). https://www.kosovopolice.com/

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, *40*, Article e253. https://doi.org/10.1017/S0140525X16001837

Oronowicz-Jaśkowiak, W. (2019). The application of neural networks in the work of forensic experts in child abuse cases. *Advances in Psychiatry and Neurology/Postępy Psychiatrii i Neurologii*, *28*(4), 273–282. https://doi.org/10.5114/ppn.2019.92489

Sejdiu, L., Shala, F., Tollazzi, T., & Demolli, H. (2024). Analysis of traffic safety factors and their impact using machine learning algorithms. *Civil Engineering Journal*, *10*(9). https://doi.org/10.28991/CEJ-2024-010-09-06

Shala, F., Demolli, H., & Sejdiu, L. (2024). The application of machine learning algorithms in predicting traffic accidents by considering weather conditions. *International Review of Civil Engineering (IRECE)*, *15*(2), 201–208. https://doi.org/10.15866/irece.v15i2.24368

Statistic Road Accident. (2024). https://statystyka.policja.pl/

Statistics Poland. (2022). Statistical Yearbook of the Republic of Poland 2022. Warsaw, Poland: Statistics Poland. https://stat.gov.pl/en/topics/statistical-yearbooks/statistical-yearbooks/statistical-yearbook-of-the-republic-of-poland-2022%2C2%2C24.html

Vilaça, M., Silva, N., & Coelho, M. C. (2017). Statistical analysis of the occurrence and severity of crashes involving vulnerable road users. *Transportation Research Procedia, 27*, 1113–1120. https://doi.org/10.1016/j.trpro.2017.12.068.

WBRSO. (2023). Western Balkans Road Safety Observatory. https://www.transport-community.org/wbrso/

Wu, Y. et al. (2020). Google's neural machine translation system: bridging the gap between human and machine translation. *arXiv:1609.08144*. https://doi.org/10.48550/arXiv.1609.08144