BEZPIECZEŃSTWO SECURITY
TEORIA I PRAKTYKA THEORY AND PRACTICE
e-ISSN 2451-0718 2024
ISSN 1899-6264 Nr 3 (LVI)

DOI: 10.48269/2451-0718-btip-2024-3-009
Data wpłynięcia: 8.01.2024
Data akceptacji: 25.06.2024

Piotr Gorzelańczyk

BEng, PhD, Associate Professor, Stanisław Staszic State University of Applied Sciences in Piła
https://orcid.org/0000-0001-9662-400X

Janusz Drzewiecki

BEng, MSc, Stanisław Staszic State University of Applied Sciences in Piła
https://orcid.org/0009-0009-0173-0981

Employing neural networks to predict the number of incidents on specific types of Polish roads

Literature review

Road accidents are situations that result in both property damage and injury or death to other motorists. Every year, 1.3 million people die in car accidents, according to the WHO. Road accidents cause a 3% GDP loss in the typical nation worldwide. Children and adults between the ages of 5 and 29 die most frequently from traffic accidents[1]. By 2030, the UN General Assembly wants to see a 50% decrease in traffic accident fatalities and injuries.

A traffic collision's size is a factor in evaluating how serious it is. In order for the responsible authorities to develop road safety legislation with the intention of preventing accidents, minimizing injuries, fatalities, and property damage, it is critical to quantify the severity of accident.[2] Before implementing countermeasures to

[1] WHO Team, *The global status on road safety*, https://www.who.int/publications/i/item/9789241565684 [accessed: 2.02.2024].

[2] T. Tambouratzis *et al.*, *Maximising accuracy and efficiency of traffic accident prediction combining information mining with computational intelligence approaches and decision trees*, "Jour-

prevent and minimize accident severity, it is vital to identify the critical components that influence accident severity.[3] A multi-node Deep Neural Network (DNN) architecture is provided by Yang et al. for forecasting various degrees of injury, fatalities, and property loss. It makes it possible to fully and accurately assess the seriousness of traffic accidents.[4]

The accident numbers come from a number of sources. Typically, government officials use the pertinent government agencies to obtain and evaluate them. Numerous sources, including hospital files, insurance company databases, and police reports, are used to collect data. As a result, the transportation industry is conducting more extensive analyses of data related to traffic accidents.[5]

Currently, the most significant information source for the analysis and forecasting of traffic events is intelligent transportation systems. GPS equipment mounted on moving vehicles may be used to analyze this data.[6] Roadside microwave vehicle detection systems may continually capture information about moving vehicles, such as speed, traffic volume, and vehicle type.[7] Additionally, a lot of traffic data may be gathered over a predetermined period of time using a license plate recognition system.[8] Social media is another possible source for information on traffic and accidents, albeit the accuracy of the reports may not be sufficient owing to the inexperience of the reporters.[9]

---

nal of Artificial Intelligence and Soft Computing Research" 2014, vol. 4, no. 1, pp. 31–42, https://doi.org/10.2478/jaiscr-2014-0023; L. Zhu *et al.*, *Analysis of accident severity for curved roadways based on Bayesian networks*, "Sustainability" 2019, vol. 11, no. 8, 2223, https://doi.org/10.3390/su11082223.

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

4 Z. Yang, W. Zhang, J. Feng, *Predicting multiple types of traffic accident severity with explanations: a multi-task deep learning framework*, "Safety Science", 2022, vol. 146, 105522, https://doi.org/10.1016/j.ssci.2021.105522.

5 P. Gorzelańczyk, D. Pyszewska, T. Kalina, M. Jurkovič, *Analysis of road traffic safety in the Piła Poviat*, "Scientific Journal of Silesian University of Technology. Series Transport" 2020, vol. 107, pp. 33–52, https://doi.org/10.20858/sjsutst.2020.107.3.

6 C. Chen, *Analysis and forecast of traffic accident big data*, "ITM Web of Conferences" 2017, vol. 12, 04029, https://doi.org/10.1051/itmconf/20171204029.

7 K.A. Khaliq *et al.*, *Road accidents detection, data collection and data analysis using V2X communication and edge/cloud computing*, "Electronics" 2019, vol. 8, no. 8, 896, https://doi.org/10.3390/electronics8080896.

8 H. Rajput, T. Som, S. Kar, *An automated vehicle license plate recognition system*, "Computer" 2015, vol. 48, no. 8, pp. 56–61, https://doi.org/10.1109/MC.2015.244.

9 Z. Zheng *et al.*, *Framework for fusing traffic information from social and physical transportation data*, "PLoS One" 2018, vol. 13, no. 8, e0201531, https://doi.org/10.1371/journal.pone.0201531.

Work with various data sources that need to be properly questioned in order for accident data to be of any value. Analytical outcomes can be more precise by integrating many data sources and combining diverse traffic accident data.[10]

To determine the severity of the issue and establish a connection between traffic participants and accidents, Vilaça et al.[11] carried out a statistical analysis. The study's findings include raising the bar for traffic safety rules and implementing more traffic safety precautions.

Based on the quantity of traffic accidents, which serves as a barometer for the investigation into accident causes, Bąk et al.[12] conducted a statistical analysis of traffic safety in a chosen Polish area. The study examined the safety variables of persons who cause accidents using multivariate statistical analysis.

The type of traffic problem being addressed determines the source of accident data to be used for analysis. The accuracy of accident prediction and accident elimination is increased by combining statistical models with additional data from real driving or other information collected from intelligent traffic systems.[13]

The literature has a number of techniques for predicting the probability of accidents. The most popular methodologies for forecasting accident frequency[14] have the drawback of not allowing for the evaluation of prediction accuracy based on previous forecasts and the frequent residual component of autocorrelation.[15] While Sunny et al.[16] employed the Holt-Winters exponential smoothing approach,

---

10 E. Abdullah, A. Emam, *Traffic accidents analyzer using big data*, [in:] *Proceedings of 2015 International Conference on Computational Science and Computational Intelligence*, eds. H.R. Arabnia, L. Deligiannidis, Q.-N. Tran, IEEE, 2015, https://doi.org/10.1109/CSCI.2015.187.

11 M. Vilaça, N. Silva, M.C. Coelho, *Statistical analysis of the occurrence and severity of crashes involving vulnerable road users*, "Transportation Research Procedia" 2017, vol. 27, p. 1113–1120, https://doi.org/10.1016/j.trpro.2017.12.113.

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

13 A. Chand, S. Jayesh, A.B. Bhasi, *Road traffic accidents: an overview of data sources, analysis techniques and contributing factors*, "Materials Today: Proceedings" 2021, vol. 47, no. 15, pp. 5135–5141, https://doi.org/10.1016/j.matpr.2021.05.415.

14 A.F. Helgason, *Fractional integration methods and short time series: evidence from a simulation study*, "Political Analysis" 2016, vol. 24, no. 1, pp. 59–68, https://doi.org/10.1093/pan/mpv026; S. Lavrenz *et al.*, *Time series modeling in traffic safety research*, "Accident Analysis and Prevention" 2018, vol. 117, pp. 368–380, https://doi.org/10.1016/j.aap.2017.11.030.

15 [L. Kowalski], *Prognozowanie na podstawie szeregów czasowych*, http://pis.rezolwenta.eu.org/Materialy/PiS-W-5.pdf [accessed: 1.02.2024].

16 C.M. Sunny et al., *Forecasting of road accident in Kerala: a case study*, [in:] *2018 International Conference on Data Science and Engineering (ICDSE)*, IEEE, 2018, https://doi.org/10.1109/ICDSE.2018.8527825.

Procházka et al.[17] used a multi-seasonality model. One of the model's drawbacks is that exogenous variables cannot be included.[18]

The frequency of traffic accidents has been predicted using the curve-fitting regression models of Al-Madani[19] and Monedero et al. for analyzing the number of fatalities,[20] as well as the vector autoregressive model, which has the disadvantage of requiring many observations of variables to accurately estimate their parameters.[21] Assuming the series are already stationary, these just require an order of autoregression[22] and a few straightforward linear connections.[23]

Random Forest regression was used by Biswas et al.[24] to forecast the frequency of traffic accidents. The approach and peak prediction are unstable,[25] the data comprise groups with related features that are as important to the original data, and smaller groups are preferred over bigger ones in this case.[26] For the proposed forecasting problem, Chudy-Laskowska and Pisula[27] employed an autoregressive quad-

17 J. Procházka, M. Čamaj, *Modelling the number of road accidents of uninsured drivers and their severity*, [in:] *International Academic Conferences*, 5408040, International Institute of Social and Economic Sciences, 2017, https://ideas.repec.org/p/sek/iacpro/5408040.html [accessed: 1.02.2024].

18 M. Szmuksta-Zawadzka, J. Zawadzki, *O prognozowaniu na podstawie modeli Holta-Wintersa dla pełnych i niepełnych danych*, "Ekonometria" 2009, vol. 24, no. 38, pp. 85–99, https://www.dbc.wroc.pl/dlibra/doccontent?id=15648 [accessed: 1.02.2024].

19 H.M.N. Al-Madani, *Global road fatality trends' estimations based on country-wise micro level data*, "Accident Analysis and Prevention" 2018, vol. 111, pp. 297–310, https://doi.org/10.1016/j.aap.2017.11.035.

20 B.D. Monedero, L.A. Gil-Alana, M.C.V. Martinez, *Road accidents in Spain: Are they persistent?*, "IATSS Research" 2021, vol. 45, no. 3, pp. 317–325, https://doi.org/10.1016/j.iatssr.2021.01.002.

21 A. Wójcik, *Modele wektorowo-autoregresyjne jako odpowiedz na krytykę strukturalnych wielorównaniowych modeli ekonometrycznych*, "Studia Economiczne" 2014, vol. 193, pp. 112–128, https://cejsh.icm.edu.pl/cejsh/element/bwmeta1.element.desklight-538707d9-40cd-471b-a022-6190a01eb76f [accessed: 2.02.2024].

22 M. Piłatowska, *Wybór rzędu autoregresji w zależności od parametrów modelu generującego*, "Ekonometria" 2012, vol. 4, no. 38, pp. 16–35, https://dbc.wroc.pl/Content/22753/Pilatowska_Wybor_Rzedu_Autoregresji_w_Zale%C5%BCnosci_Od_Parametrow.pdf [accessed: 1.02.2024].

23 M. Mamczur, *Jak działa regresja liniowa? I czy warto ją stosować?*, https://miroslawmamczur.pl/jak-dziala-regresja-liniowa-i-czy-warto-ja-stosowac/ [accessed: 2.02.2024].

24 A.A. Biswas, J. Mia, A. Majumder, *Forecasting the number of road accidents and casualties using random forest regression in the context of Bangladesh*, [in:] *Proceedings of 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, IEEE, 2019, https://doi.org/10.1109/ICCCNT45670.2019.8944500.

25 *Random forest*, https://en.wikipedia.org/wiki/Random_forest [accessed: 2.02.2024].

26 K. Fijorek, *et al.*, *Prognozowanie cen energii elektrycznej na Rynku Dnia Następnego metodami data mining*, "Rynek Energii" 2010, vol. 91, no. 6, pp. 46–50, https://www.cire.pl/pliki/2/prognozowanie_cen_rdn.pdf [accessed: 2.02.2024].

27 K. Chudy-Laskowska, T. Pisula, *Prognozowanie liczby wypadków drogowych na Podkarpaciu*, „Logistyka" 2015, no. 4, pp. 2782–2796.

ratic trend model, a univariate periodic trend model, and an exponential smoothing model. The problem at hand may potentially be anticipated using a moving average model, however this approach has poor forecast accuracy, data loss within a sequence, and is unable to take into account trends and seasonal fluctuations.[28]

The GARMA approach, which restricts the parameter space, was employed by Procházka et al.[29] to ensure that the process is stable. Forecasting frequently uses the ARMA model for stationary systems [30] and the ARIMA or SARIMA model for non-stationary phenomena. The benefit of these models is that they give the models under investigation a considerable deal of flexibility; nevertheless, the drawback is that they demand more advanced research skills from the researcher than, for instance, regression analysis.[31] The linearity of the ARIMA model is another drawback.[32]

In their study[33] Chudy-Laskowska and Pisula employed an ANOVA to forecast the frequency of traffic accidents. This approach's drawback is that it makes extra assumptions, most notably the assumption of sphericity, the failure of which could result in incorrect findings.[34] The incidence of vehicle accidents is also predicted using neural network approaches. Because ANNs are frequently referred to as "black boxes," where input data is entered and the model outputs results without being aware of the analysis, they have a number of drawbacks, including the requirement

---

28 N. Kashpruk, *Badania porównawcze modeli statystycznych i obliczeń miękkich dla identyfikacji ciągów czasowych i prognoz*, PhD dissertation, Opole University of Technology, Opole 2010, https://dbc.wroc.pl/Content/108023/Praca%20Doktorska_%20Nataliia%20Kashpruk_popr.pdf [accessed: 2.02.2024].

29 J. Procházka *et al.*, *Modelling the number of road accidents*, [in:] *20th AMSE. Applications of Mathematics and Statistics in Economics. International Scientific Conference: Szklarska Poręba, 30 August – 3 September 2017. Conference Proceedings. Full Text Papers*, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław 2017, pp. 355–364, https://doi.org/10.15611/amse.2017.20.29.

30 B. Dutta, M.P. Barman, A.N. Patowary, *Application of Arima model for forecasting road accident deaths in India*, "International Journal of Agricultural and Statistical Sciences" 2020, vol. 16, no. 2, pp. 607–615, https://connectjournals.com/03899.2020.16.607; M. Karlaftis, E. Vlahogianni, *Memory properties and fractional integration in transportation time-series*, "Transportation Research Part C: Emerging Technologies" 2009, vol. 17, no. 4, pp. 444–453, https://doi.org/10.1016/j.trc.2009.03.001.

31 S. Łobejko, K. Masłowska, R. Wojdan, *Analiza i prognozowanie szeregów czasowych z programem SAS*, Oficyna Wydawnicza SGH, Warszawa 2015.

32 M. Szmuksta-Zawadzka, J. Zawadzki, *op. cit*.

33 K. Chudy-Laskowska, T. Pisula, *Prognoza liczby wypadków drogowych w Polsce*, "Logistyka" 2014, no. 6, p. 2710–2722.

34 A. Gregorczyk, M. Swarcewicz, *Analiza wariancji w układzie powtarzanych pomiarów do określenia efektów czynników wpływających na pozostałości linuronu w glebie*, "Polish Journal of Agronomy" 2012, vol. 11, pp. 15–20, https://www.iung.pl/PJA/wydane/11/PJA11_3.pdf [accessed: 2.02.2024].

for prior expertise in this field[35], the dependence of the final result on the network's initial conditions, and the impossibility of conventionally interpreting results.[36]

Kumar et al.[37] used the Hadoop model as a cutting-edge prediction technique. This strategy's drawback is that it is unable to handle tiny data sets.[38] The GARCH model was employed by Karlaftis and Vlahogianni[39] to provide predictions. This strategy's intricate model and complicated form are a problem.[40] However, McIlroy[41] and his team's usage of the ADF test has the drawback of not having sufficient power to detect the autocorrelation of the random component.[42]

The authors created predictions for the number of accidents on Polish roads based on the aforementioned data. In order to predict the amount of accidents, neural networks were deployed.

Materials and methods

On Polish roadways, several incidents happen every year. The epidemic has decreased the amount of auto accidents in recent years, which has an impact on the received prediction value. The number of traffic accidents is still too high despite the epidemic. Due to this, every effort should be made to lower this number and identify the kinds of routes where the majority of accidents will occur (Figure 1). On highways with one carriageway running in both directions, the majority of traffic accidents take place.

---

35 K. Chudy-Laskowska, T. Pisula, *Prognoza liczby wypadków drogowych w Polsce*, *op. cit.*; M.S. Wróbel, *Zastosowanie neuronowych systemów rozmytych w chemii*, praca doktorska, Politechnika Śląska, Katowice 2011, https://rebus.us.edu.pl/bitstream/20.500.12128/5266/1/Wrobel_Zastosowanie_neuronowych_systemow.pdf [accessed: 2.02.2024].

36 *Techniki zgłębiania danych (data mining)*, StatSoft, https://www.statsoft.pl/textbook/stathome_stat.html?https%3A%2F%2Fwww.statsoft.pl%2Ftextbook%2Fstdatmin.html [accessed: 2.02.2024].

37 S. Kumar, V. Viswanadham, B. Bharathi, *Analysis of road accident*, "IOP Conference Series: Materials Science and Engineering" 2019, vol. 590, no. 1, 012029, https://doi.org/10.1088/1757-899X/590/1/012029.

38 *Top advantages and disadvantages of Hadoop 3*, DataFlair, https://data-flair.training/blogs/advantages-and-disadvantages-of-hadoop/ [accessed: 2.02.2024].

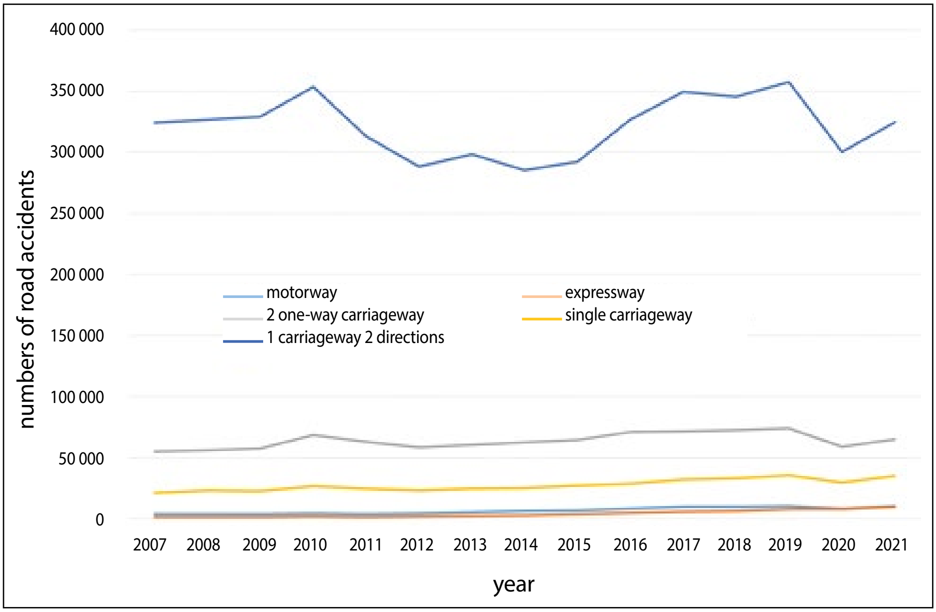
39 S. Łobejko, K. Masłowska, R. Wojdan, *op. cit*.

40 G. Perczak, P. Fiszeder, *Model GARCH – wykorzystanie dodatkowych informacji o cenach minimalnych i maksymalnych*, "Bank i Kredyt" 2014, vol. 45, no. 2, pp. 105–131, https://www.bankandcredit.nbp.pl/content/2014/02/bik_02_2014_02_art.pdf [accessed: 2.02.2024].

41 R.C. McIlroy, *et al.*, *Who is responsible for global road safety? A cross-cultural comparison of Actor Maps*, "Accident Analysis and Prevention" 2019, vol. 122, pp. 8–18, https://doi.org/10.1016/j.aap.2018.09.011.

42 J. Mućk, *Ekonometria. Modelowanie szeregów czasowych. Stacjonarnośc. Testy pierwiastka jednostkowego. Modele ARDL*, http://web.sgh.waw.pl/~jmuck/Ekonometria/EkonometriaPrezentacja5.pdf [accessed: 2.02.2024].

Figure 1. Number of road accidents in Poland by road type in 2007–2021



Source: authors' own elaboration based on: *Wypadki drogowe – statystyki roczne*, https://statystyka.policja.pl/st/ruch-drogowy/76562,Wypadki-drogowe-raporty-roczne.html [accessed: 02.02.2024].

Depending on the kind of road, specific neural network models were employed to forecast how many accidents will occur on Polish roads. This approach has the benefit of simulating how the human brain functions. Nodes make up a neural network, and each node has inputs, weights, variances, and outputs. The Statistica program chose the best weights for the investigation. The model and its parameters that are selected will determine the outcome of predicting using the approach covered above.

Measures of analytical forecasting brilliance were computed using the forecasting errors derived from equations as follows:

- ME – mean error

$$ME = \frac{1}{n} \Sigma_{i=1}^{n} (Y_i - Y_p) \tag{1}$$

- MAE – mean everage error

$$MAE = \frac{1}{n} \Sigma_{i=1}^{n} |Y_i - Y_p| \tag{2}$$

- MPE – mean percentage error

$$MPE = \frac{1}{n} \Sigma_{i=1}^{n} \frac{Y_i - Y_p}{Y_i} \tag{3}$$

- MAPE – mean absolute percentage error

$$MAPE = \frac{1}{n} \Sigma_{i=1}^{n} \frac{|Y_i - Y_p|}{Y_i} \tag{4}$$

- MSE – mean square error

$$MSE = \frac{1}{n} \Sigma_{i=1}^{n} (Y_i - Y_p)^2 \tag{5}$$

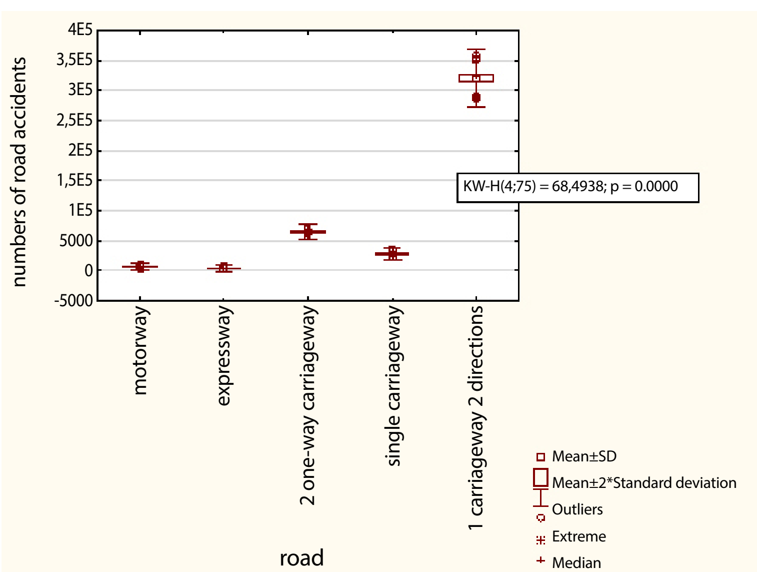
where:
n – the length of the forecast horizon,
Y – observed value of road accidents,
Y_p – forecasted value of road accidents.

The number of traffic accidents by type of road was predicted using the neural network models with the fewest mean absolute percentage error and mean percentage error.

Results

A statistical test was conducted to compare the variations in the number of traffic accidents across the studied period. With a test probability of p=0.05, the statistic for the examined non-parametric Kruskal-Wallis test is 68.5. As a result, we can rule out the idea that the average number of traffic accidents across all types of roads was similar over the study period (Figure 2).

Figure 2. From 2007 to 2021, the average number of accidents on Poland's roads was compared by kind of road



Source: *Wypadki drogowe – statystyki roczne*, https://statystyka.policja.pl/st/ruch-drogowy/76562,Wypadki-drogowe-raporty-roczne.html [accessed: 02.02.2024].

The yearly number of traffic accidents in Poland was predicted based on the type of route using information from the Polish Police from 2007 to 2021[43]. The research was carried out using Statistica software, using two different random sample sizes:

- teaching 70%, testing 15% and validation 15%.
- teaching 80%, testing 10% and validation 10%,

with the following number of learning networks: 20, 40, 60, 80, 100, 200, for which the MP error value was minimal (Table 1 and 2).

Table 1. Neural network learning summary for the random sample size case teaching 70%, testing 15% and validation 15%

Type of road	Network number	Network name	Quality (learning)	Quality (learning)	Quality (validation)	Learning algorithm	Activation (hidden)	Activation (output)	Errors				
									ME	MAE	MPE	MAPE	SSE
motorway	200	MLP 4-4-1	0,860209	0,00	1,000000	BFGS 4	Tanh	Exponential	148,860757	646,2109	0,14%	9,26%	829,6835
expressway	200	MLP 4-5-1	0,943807	0,00	1,000000	BFGS 0	Tanh	Tanh	119,16107	439,3001	0,55%	9,25%	579,3047595
2 single carriageways	80	MLP 4-7-1	0,688013	0,00	1,00000	BFGS 7	Exponential	Logistics	357,574337	2373,703	0,66%	3,69%	2900,535546
one-way	80	MLP 4-4-1	0,889765	0,00	-1,00000	BFGS 5	Linear	Tanh	240,536048	1437,768	0,88%	5,15%	2147,209989
1 carriageway 2 directions	60	MLP 4-7-1	0,769253	0,00	1,000000	BFGS 6	Logistics	Tanh	428,091314	13255,06	0,15%	4,18%	16147,18779

Source: authors' own elaboration.

Table 2. Neural network learning summary for the random sample size case teaching 80%, testing 10% and validation 10%

Type of road	Network number	Network name	Quality (learning)	Quality (learning)	Quality (validation)	Learning algorithm	Activation (hidden)	Activation (output)	Errors				
									ME	MAE	MPE	MAPE	SSE
motorway	20	MLP 4-8-1	0,896937	0,00	0,00	BFGS 3	Logistics	Linear	286,043436	830,1927	0,69%	11,54%	915,6960065
expressway	40	MLP 4-4-1	0,920674	0,00	0,00	BFGS 5	Logistics	Tanh	27,3902845	469,6358	0,93%	8,95%	686,2141408
2 single carriageways	60	MLP 4-4-1	0,823026	0,00	0,00	BFGS 18	Exponential	Logistics	332,171446	2193,8	0,51%	3,42%	2526,060222
one-way	20	MLP 4-7-1	0,874992	0,00	0,00	BFGS 1	Exponential	Logistics	540,759644	3736,17	0,27%	12,80%	4305,599907
1 carriageway 2 directions	20	MLP 4-5-1	0,912915	0,00	0,00	BFGS 8	Logistics	Linear	122,435598	6012,833	0,03%	1,90%	8468,116816

Source: authors' own elaboration.

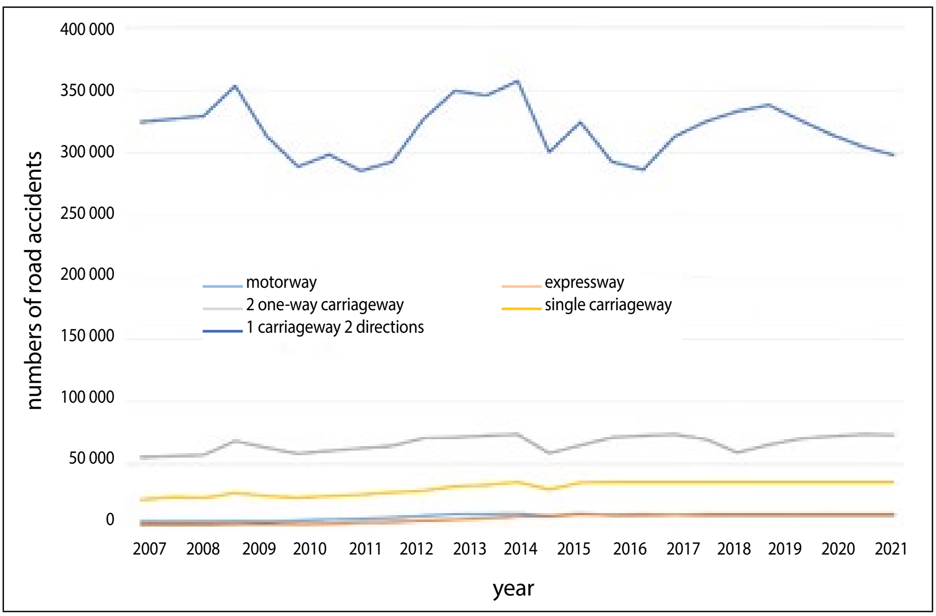
The offered research findings allow for the conclusion that, regardless of the kind of route, similar number of accidents will occur on Polish roads each year.

[43] *Wypadki drogowe – statystyki roczne*, https://statystyka.policja.pl/ [accessed: 2.02.2024].

The choice of the random sample size affects the outcomes. The average percentage error is reduced by increasing the proportion of the learning group in comparison to the test and validation group. The error was 2.28% for a learning group made up of 70%, a test group made up of 15%, and a validation group made up of 15% in the ratio (70-15-15), whereas the error for the second test (80-10-10) was 0.55% (Figures 3 and 4).

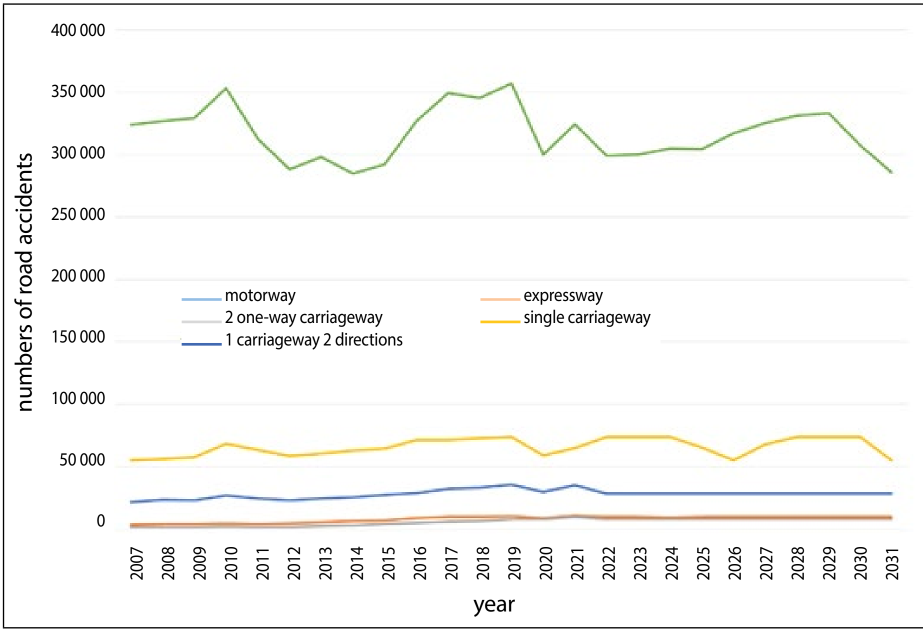
We may still anticipate a stabilization of the number of road accidents based on the data acquired on the expected number of accidents on various types of roads. This is impacted by the rise in traffic on Polish roads and the construction of new highways. It should be mentioned that the epidemic had an impact on the outcomes.

Figure 3. Forecasting number of road accidents for 2022–2040 for the 70-15-15 test group



Source: authors' own elaboration.

Figure 4. Forecasting number of road accidents for 2022–2040 for the 80-10-10 test group



Source: authors' own elaboration.

Conclusions

The study was conducted out in the Statistica environment and employed neural networks to forecast the frequency of accidents in Poland according to road type. The computer evaluated the weights used in the research to minimize the mean absolute error and the mean absolute percentage error.

It is clear from the data that we may still anticipate a stabilization in the number of traffic accidents. This is impacted by the expansion of highways and by the rise in the number of cars on Polish roads. The estimated forecasting errors demonstrate the validity of the used models.

Actions should be done to further reduce the number of road accidents based on the derived projections. These changes may include introducing harsher fines for traffic violations. The epidemic, which significantly reduced the frequency of accidents on the road, undoubtedly had an impact on the findings of the study.

In their follow-up study, the authors want to take into account more variables that affect the frequency of accidents and employ other statistical techniques to calculate the number of collisions. These might include the amount of traffic, the kind of weather, the age of the accident's perpetrator, and approaches that use exponential growth to calculate the frequency of accidents.

References

Abdullah E., Emam A., *Traffic accidents analyzer using big data*, [in:] *Proceedings of 2015 International Conference on Computational Science and Computational Intelligence*, eds. H.R. Arabnia, L. Deligiannidis, Q.-N. Tran, IEEE, 2015, pp. 392–397, https://doi.org/10.1109/CSCI.2015.187.

Al-Madani H.M.N., *Global road fatality trends' estimations based on country-wise micro level data*, "Accident Analysis and Prevention" 2018, vol. 111, pp. 297–310, https://doi.org/10.1016/j.aap.2017.11.035.

Arteaga C., Paz A., Park J., *Injury severity on traffic crashes: a text mining with an interpretable machine-learning approach*, "Safety Science" 2020, vol. 132, 104988.

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

Biswas A.A., Mia J., Majumder A., *Forecasting the number of road accidents and casualties using random forest regression in the context of Bangladesh*, [in:] *Proceedings of 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, IEEE, 2019, https://doi.org/10.1109/ICCCNT45670.2019.8944500.

Chand A., Jayesh S., Bhasi A.B., *Road traffic accidents: an overview of data sources, analysis techniques and contributing factors*, "Materials Today: Proceedings" 2021, vol. 47, no. 15, pp. 5135–5141, https://doi.org/10.1016/j.matpr.2021.05.415.

Chen C., *Analysis and forecast of traffic accident big data*, "ITM Web of Conferences" 2017, vol. 12, 04029, https://doi.org/10.1051/itmconf/20171204029.

Chudy-Laskowska K., Pisula T., *Prognozowanie liczby wypadków drogowych na Podkarpaciu*, „Logistyka" 2015, no. 4, pp. 2782–2796.

Chudy-Laskowska K., Pisula T., *Prognoza liczby wypadków drogowych w Polsce*, „Logistyka" 2014, no. 6, pp. 2710–2722.

Dutta B., Barman M.P., Patowary A.N., *Application of Arima model for forecasting road accident deaths in India*, "International Journal of Agricultural and Statistical Sciences" 2020, vol. 16, no. 2, pp. 607–615, https://connectjournals.com/03899.2020.16.607.

Fijorek K., Mróz K., Niedziela K., Fijorek D., *Prognozowanie cen energii elektrycznej na Rynku Dnia Następnego metodami data mining*, „Rynek Energii" 2010, vol. 91, no. 6, pp. 46–50, https://www.cire.pl/pliki/2/prognozowanie_cen_rdn.pdf [accessed: 02.02.2024].

Gorzelańczyk P., Pyszewska D., Kalina T., Jurkovič M., *Analysis of road traffic safety in the Piła Poviat*, "Scientific Journal of Silesian University of Technology. Series Transport" 2020, vol. 107, pp. 33–52, https://doi.org/10.20858/sjsutst.2020.107.3.

Gregorczyk A., Swarcewicz M., *Analiza wariancji w układzie powtarzanych pomiarów do określenia efektów czynników wpływających na pozostałości linuronu w glebie*, "Polish Journal of Agronomy" 2012, vol. 11, pp. 15–20, https://www.iung.pl/PJA/wydane/11/PJA11_3.pdf [accessed: 02.02.2024].

Helgason A.F., *Fractional integration methods and short time series: evidence from a simulation study*, "Political Analysis" 2016, vol. 24, no. 1, pp. 59–68, https://doi.org/10.1093/pan/mpv026.

Karlaftis M., Vlahogianni E., *Memory properties and fractional integration in transportation time-series*, "Transportation Research Part C: Emerging Technologies" 2009, vol. 17, no. 4, pp. 444–453, https://doi.org/10.1016/j.trc.2009.03.001.

Kashpruk N., *Badania porównawcze modeli statystycznych i obliczeń miękkich dla identyfikacji ciągów czasowych i prognoz*, PhD dissertation, Opole University of Technology, Opole 2010, https://dbc.wroc.pl/Content/108023/Praca%20Doktorska_%20Nataliia%20Kashpruk_popr.pdf [accessed: 02.02.2024].

Khaliq K.A., Chughtai O., Shahwani A., Qayyum A., Pannek J., *Road accidents detection, data collection and data analysis using V2X communication and edge/cloud computing*, "Electronics" 2019, vol. 8, no. 8, 896, https://doi.org/10.3390/electronics8080896.

[Kowalski L.], *Prognozowanie na podstawie szeregów czasowych*, http://pis.rezolwenta.eu.org/Materialy/PiS-W-5.pdf [accessed 02.02.2024].

Kumar S., Viswanadham V., Bharathi B., *Analysis of road accident*, "IOP Conference Series: Materials Science and Engineering" 2019, vol. 590, no. 1, 012029, https://doi.org/10.1088/1757-899X/590/1/012029.

Lavrenz S., Vlahogianni E., Gkritza K., Ke Y., *Time series modeling in traffic safety research*, "Accident Analysis and Prevention" 2018, vol. 117, pp. 368–380, https://doi.org/10.1016/j.aap.2017.11.030.

Łobejko S., Masłowska K., Wojdan R., *Analiza i prognozowanie szeregów czasowych z programem SAS*, Oficyna Wydawnicza SGH, Warszawa 2015.

Mamczur M., *Jak działa regresja liniowa? I czy warto ja stosować?*, https://miroslawmamczur.pl/jak-dziala-regresja-liniowa-i-czy-warto-ja-stosowac/ [accessed 02.02.2024].

McIlroy R.C., Plant K.A., Hoque M.S., Wu J., Kokwaro G.O., Nam V.H., Stanton N.A., *Who is responsible for global road safety? A cross-cultural comparison of Actor Maps*, "Accident Analysis and Prevention" 2019, vol. 122, pp. 8–18, https://doi.org/10.1016/j.aap.2018.09.011.

Monedero B.D., Gil-Alana L.A., Martinez M.C.V., *Road accidents in Spain: Are they persistent?*, "IATSS Research" 2021, vol. 45, no. 3, pp. 317–325, https://doi.org/10.1016/j.iatssr.2021.01.002.

Mućk J., *Ekonometria. Modelowanie szeregów czasowych. Stacjonarność. Testy pierwiastka jednostkowego. Modele ARDL*, http://web.sgh.waw.pl/~jmuck/Ekonometria/EkonometriaPrezentacja5.pdf [accessed: 02.02.2024].

Perczak G., Fiszeder P., *Model GARCH – wykorzystanie dodatkowych informacji o cenach minimalnych i maksymalnych*, "Bank i Kredyt" 2014, vol. 45, no. 2, pp. 105–131, https://www.bankandcredit.nbp.pl/content/2014/02/bik_02_2014_02_art.pdf [accessed: 02.02.2024].

Piłatowska M., *Wybór rzędu autoregresji w zależności od parametrów modelu generującego*, „Ekonometria" 2012, vol. 4, no. 38, pp. 16–35, https://dbc.wroc.pl/Content/22753/Pilatowska_Wybor_Rzedu_Autoregresji_w_Zale%C5%BCnosci_Od_Parametrow.pdf [accessed: 02.02.2024].

Prochazka J., Camaj M., *Modelling the number of road accidents of uninsured drivers and their severity*, [in:] *International Academic Conferences*, 5408040, International Institute of Social and Economic Sciences, 2017, https://ideas.repec.org/p/sek/iacpro/5408040.html [accessed 02.02.2024].

Procházka J., Flimmel S., Čamaj M., Bašta M., *Modelling the number of road accidents*, [in:] *20th AMSE. Applications of Mathematics and Statistics in Economics. International Scientific Conference: Szklarska Poręba, 30 August – 3 September 2017. Conference Proceedings. Full Text Papers*, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław 2017, pp. 355–364, https://doi.org/10.15611/amse.2017.20.29.

Rajput H., Som T., Kar S., *An automated vehicle license plate recognition system*, "Computer" 2015, vol. 48, no. 8, pp. 56–61, https://doi.org/10.1109/MC.2015.244.

Random forest, https://en.wikipedia.org/wiki/Random_forest [accessed: 02.02.2024].

Sunny C.M., Nithya S., Sinshi K.S., Vinodini V.M.D., Lakshmi A.K.G., Anjana S., Manojkumar T.K., *Forecasting of road accident in Kerala: a case study*, [in:] *2018 International*

Conference on Data Science and Engineering (ICDSE), IEEE, 2018, https://doi.org/10.1109/ICDSE.2018.8527825.

Szmuksta-Zawadzka M., Zawadzki J., *O prognozowaniu na podstawie modeli Holta-Wintersa dla pełnych i niepełnych danych*, "Ekonometria" 2009, vol. 24, no. 38, pp. 85–99, https://www.dbc.wroc.pl/dlibra/doccontent?id=15648 [accessed: 02.02.2024].

Tambouratzis T., Souliou D., Chalikias M., Gregoriades A., *Maximising accuracy and efficiency of traffic accident prediction combining information mining with computational intelligence approaches and decision trees*, "Journal of Artificial Intelligence and Soft Computing Research" 2014, vol. 4, no. 1, pp. 31–42, https://doi.org/10.2478/jaiscr-2014-0023.

Techniki zgłębiania danych (data mining), StatSoft, https://www.statsoft.pl/textbook/stathome_stat.html?https%3A%2F%2Fwww.statsoft.pl%2Ftextbook%2Fstdatmin.html [accessed: 02.02.2024].

Top advantages and disadvantages of Hadoop 3, DataFlair, https://data-flair.training/blogs/advantages-and-disadvantages-of-hadoop/ [accessed: 02.02.2024].

Vilaça M., Silva N., Coelho M.C., *Statistical analysis of the occurrence and severity of crashes involving vulnerable road users*, "Transportation Research Procedia" 2017, vol. 27, pp. 1113–1120, https://doi.org/10.1016/j.trpro.2017.12.113.

WHO Team, *The global status on road safety*, https://www.who.int/publications/i/item/9789241565684 [accessed: 02.02.2024].

Wójcik A., *Modele wektorowo-autoregresyjne jako odpowiedz na krytykę strukturalnych wielorównaniowych modeli ekonometrycznych*, „Studia Ekonomiczne" 2014, vol. 193, pp. 112–128, https://cejsh.icm.edu.pl/cejsh/element/bwmeta1.element.desklight-538707d9-40cd-471b-a022-6190a01eb76f [accessed 02.02.2024].

Wróbel M.S., *Zastosowanie neuronowych systemów rozmytych w chemii*, praca doktorska, Politechnika Śląska, Katowice 2011, https://rebus.us.edu.pl/bitstream/20.500.12128/5266/1/Wrobel_Zastosowanie_neuronowych_systemow.pdf [accessed: 02.02.2024].

Wypadki drogowe – statystyki roczne, https://statystyka.policja.pl/st/ruch-drogowy/76562,Wypadki-drogowe-raporty-roczne.html [accessed: 02.02.2024].

Yang Z., Zhang W., Feng J., *Predicting multiple types of traffic accident severity with explanations: a multi-task deep learning framework*, "Safety Science" 2022, vol. 146, 105522 https://doi.org/10.1016/j.ssci.2021.105522.

Zheng Z., Wang C., Wang P., Xiong Y., Zhang F., Lv Y., *Framework for fusing traffic information from social and physical transportation data*, "PLoS One" 2018, vol. 13, no. 8, e0201531https://doi.org/10.1371/journal.pone.0201531.

Zhu L., Lu L., Zhang W., Zhao Y., Song M., *Analysis of accident severity for curved roadways based on Bayesian networks*. Sustainability. 2019, vol. 11, no. 8, 2223 https://doi.org/10.3390/su11082223.

Employing neural networks to predict the number of incidents on specific types of Polish roads

Abstract

The article's goal is to predict how many accidents will occur on different types of roads in Poland. This was accomplished by the analysis of annual data on the number of traffic accidents in Poland by type of road. A prediction for the years 2022–2040 was developed using police statistics. The frequency of accidents in Poland was anticipated using a few neural network models. The findings indicate that we can still expect a stabilization of the number of road accidents. This is impacted by the rise in traffic on Polish roads and the construction of new highways. The number of learning, test, and validation samples chosen at random has an impact on the outcomes.

Keywords: road accident, pandemic, forecasting, neural networks

Prognozowanie liczby wypadków drogowych w Polsce dla poszczególnych rodzajów dróg przy wykorzystaniu sieci neuronowych

Streszczenie

Celem artykułu jest wykonanie prognozy liczby wypadków drogowych w Polsce w zależności od rodzaju drogi. Dokonano analizy rocznych danych dotyczących liczby wypadków drogowych w Polsce z podziałem na kategorie dróg i na podstawie statystyk Policji sporządzono prognozę na latach 2022–2040. Do prognozy liczby wypadków w Polsce wykorzystano wybrane modele sieci neuronowych. Wyniki badań pokazują, że nadal możemy się spodziewać stabilizacji liczby wypadków drogowych. Wpływ ma na to rosnąca liczba pojazdów poruszających się na polskich drogach i tworzenie nowych dróg szybkiego ruchu i autostrad. Wybór liczności prób losowych (uczącej, testowej i walidacji) wpływa na otrzymane wyniki.

Słowa kluczowe: wypadek drogowy, pandemia, prognozowanie, sieci neuronowe