The Use of Neural Networks to Predict Frequency of Road Accidents in Poland and Slovenia

Piotr GORZELAŃCZYK[1], Marina ZANNE[2]

[1] Corresponding author, piotr.gorzelanczyk@ans.pila.pl, Department of Transport, Stanislaw Staszic State University of Applied Sciences, Piła, Poland
[2] marina.zanne@fpp.uni-lj.si, Faculty of Maritime Studies and Transport, University of Ljubljana, Portorož, Slovenia

ABSTRACT

Despite a general decline in recent years, road accidents are still a major problem worldwide, including in Poland and Slovenia. Even if there is no denying that the COVID-19 epidemic has had an impact on accident rates, the statistics nonetheless show that urgent action is required to further lower their frequency. The purpose of this study is to project how many traffic accidents will occur in Slovenia and Poland between 2024 and 2030. This was accomplished by analysing historical data on yearly accident incidence from Eurostat and the Polish Police. To create the predicted numbers, certain neural network models were used, utilising these datasets. The results indicate that the number of road accidents is likely to stabilise soon. This forecast is influenced by a number of factors, including the continued increase in car ownership and the ongoing development of road infrastructure, including the construction of new motorways and roads. It is important to remember that the accuracy of the forecast is susceptible to inherent limitations due to the random sampling of the data used to test, validate and train the models.

KEYWORDS

traffic accident; pandemic; forecasting; neural networks; Poland; Slovenia.

1. INTRODUCTION

Traffic accidents are a major health and financial burden worldwide. The World Health Organisation (WHO) estimates that 1.3 million people die each year from traffic-related injuries, and these incidents cost countries around the world 3% of their average GDP [1]. It is noteworthy that the leading cause of mortality among young people aged 5 to 29 years is motor vehicle accidents [1]. By 2030, the UN General Assembly has set an ambitious target to reduce the number of deaths and injuries from road traffic accidents worldwide.

To effectively develop a road safety policy, it is necessary to accurately measure the severity of traffic incidents. Implementing countermeasures to avoid and reduce the consequences of accidents requires an understanding of the elements that contribute to accident severity [2–4]. A Deep Neural Network (DNN) architecture was presented by Yang et al. [5] to predict different levels of injury, mortality and property damage, enabling accurate and precise investigation of accident severity.

Accurate analysis and accident predictions depend on reliable data sources. In the past, government organisations have collected and analysed accident data from a variety of sources such as police reports, insurance databases and medical records [6].

Data sources for traffic analysis are numerous in modern transport networks. Data on vehicle movements, speeds and traffic volumes can be continuously recorded by cars equipped with GPS, roadside microwave sensors and licence plate recognition software [7–9]. Although real-time traffic information can be obtained via social media, user-generated data is not always accurate [10].

Integration of data from several sources is often necessary for effective accident analysis. The accuracy of analytical conclusions can be improved by harmonising heterogeneous accident data and combining data from several sources [11].

Vilaca et al [12] conducted a statistical study of the relationship between road traffic participants and accident severity, highlighting the need for stricter road safety laws and regulations. Using accident data as a stand-in to investigate the causes of accidents, Bak et al [13] used multivariate statistical analysis to investigate the safety characteristics of accident perpetrators in a specific region of Poland.

The specific research question determines which data source is used for accident analysis. The accuracy of accident prediction and mitigation can be improved by combining statistical models with data from intelligent transport systems and actual driving [14].

Time series models are often used to predict accident frequency [15, 16]. However, these models can inadequately assess the accuracy of predictions based on historical forecasts and often exhibit residual autocorrelation [17]. Although some studies have used Holt-Winters exponential smoothing [19] and multi-seasonal models [18], these techniques may not adequately account for exogenous factors [20].

One of the most important measures of road safety is the number of accidents per 10,000 people (NRA). With a population of 37.6 million, Poland had 20,936 traffic accidents in 2023, yielding an NRA of 5.57. In comparison, Slovenia, which has a population of 3.86 million, had 10,633 traffic accidents, resulting in an NRA of 27.55, which is far higher than Poland's.

$$NRA = \frac{NR}{NI} * 10000$$

(1)

where:
— NR – number road accidents
— NI – number inhabitants

Using previously provided statistics, the authors estimated the number of accidents on the roads in Poland and Slovenia. Neural networks were used to predict the number of accidents in both countries.

2. MATERIALS AND METHODS

Road accidents continue to be a serious public safety problem. Although mobility restrictions caused by the pandemic have temporarily reduced the number of accidents in recent years, the total number is still too high. Implementing techniques to reduce the frequency of accidents and identifying the types of roads with the highest accident rates is essential to properly reduce this risk. Targeted action to improve road safety will be possible through a data-driven strategy.
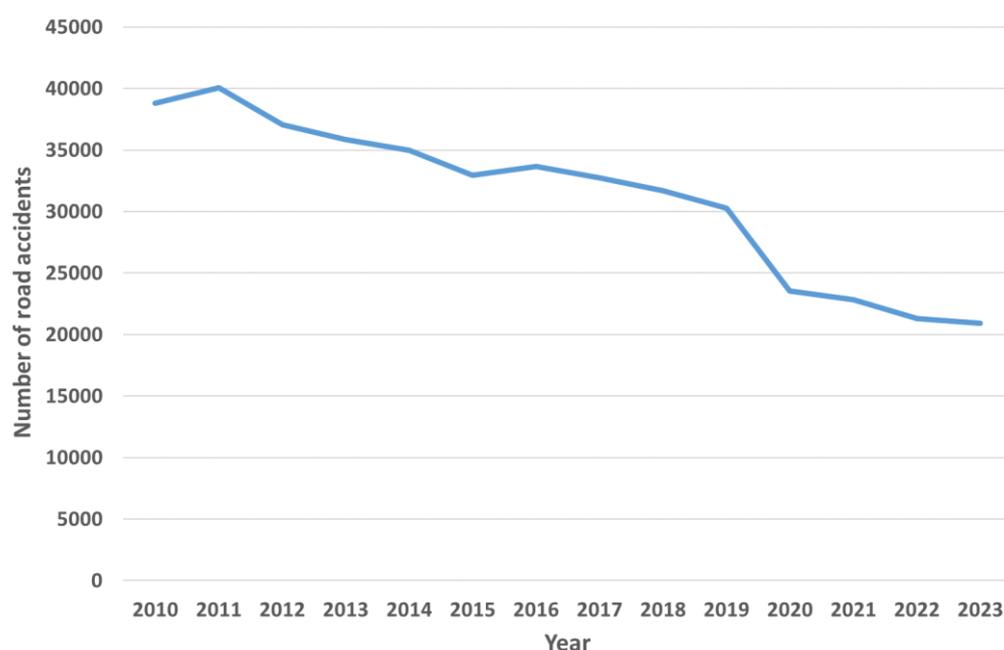

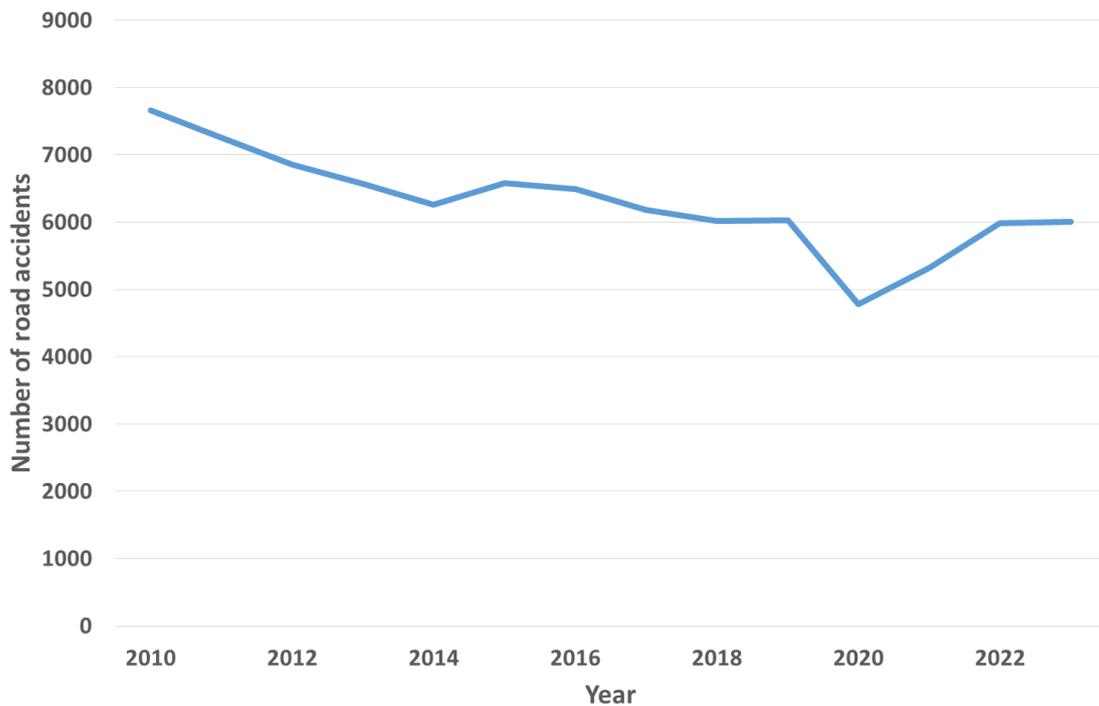*Figure 1 – Poland's traffic accident rate from 2010 to 2023 [21]*

Figure 2 – Slovenia's traffic accident rate from 2010 to 2023 [22]

To predict the number of road accidents in Poland and Slovenia, the study used special neural network models. Inspired by the human brain, neural networks consist of interconnected nodes that use a series of weighted connections to process information.

The accuracy of the prediction is strongly influenced by the choice of the best model architecture and parameters. The best weighting values for the selected model were found using the Statistica software. Because of its multilayer design, neural networks can recognise intricate patterns in incoming data, including text, voice, and pictures. To get the desired result, the network's internal parameters must be changed during the learning process. Basic building pieces of neural networks – artificial neurons – take several inputs and generate a single output, simulating the function of actual neurons.

Tailored product suggestions in online auctions, machine translation in programs like Google Translate and tailored content recommendations in streaming services are just a few of the many uses for neural networks. Neural networks were used in this study to predict how frequently traffic accidents will occur.

A multilayer perceptron neural network (MLP) was used in the prediction challenge. An input layer, one or more hidden layers (with two to eight neurons each) and an output layer formed the MLP design. Time series predictions for the number of traffic accidents were provided by the output layer [23–27].

Equations (2–7) create a set of prediction error measures that were used to assess the model's performance:

— ME – mean error

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

— MAE – mean error

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

— MPE – mean percentage error

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

— MAPE – mean absolute percentage error

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i - Y_p|}{Y_i} \tag{5}$$

— SSE – mean square error

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

— M^2 – Theila measure

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

where:
— n – length the forecast horizon,
— Y – observed value road accidents
— Y p – forecast value road accidents.

The neural network models with the lowest mean percentage error and mean absolute percentage error were employed to forecast the likelihood of traffic accidents in the future.

3. RESULTS

Polish Police statistics from 2010 to 2023 [21] and Eurostat data [22] were used to forecast the annual number of road accidents in Poland and Slovenia. Both studies were conducted using the Statistica software, assuming two random sample sizes:
1) teaching 70%, testing 15% and validation 15%.
2) teaching 80%, testing 10% and validation 10%,
with the lowest possible MP error for the following number of learning networks: 20, 40, 60, 80, 100 and 200 (*Tables 1–4*).

Table 1 – Summary of neural network learning for a random learning sample size of 70%, a test sample size of 15% and a validation sample size of 15% for Poland

Network number	Network name	Quality (learning)	Quality (testing)	Quality (validation)	Activation (hidden)	Activation (output)	Errors					
							ME	MAE	MPE	MAPE	SSE	Theil
20	MLP 1-7-1	0.97	0.98	0.99	Tanh	Logistic	932.03	2374.97	2.85%	6.69%	2747.76	4.56E-03
20	MLP 1-8-1	0.97	0.98	0.99	Linear	Logistic	664.14	2092.80	3.19%	6.46%	2667.94	4.30E-03
20	MLP 1-2-1	0.97	0.98	0.99	Exponential	Exponential	1168.74	2119.69	3.23%	5.69%	2657.54	4.26E-03
20	MLP 1-4-1	0.96	0.97	0.99	Exponential	Linear	1815.06	2891.74	2.84%	7.52%	3551.21	7.61E-03
20	MLP 1-3-1	0.97	0.97	0.99	Tanh	Logistic	1113.77	2387.70	3.25%	6.64%	2802.32	4.74E-03
40	MLP 1-7-1	0.97	0.98	0.99	Exponential	Exponential	837.43	2001.24	2.77%	5.68%	2484.15	3.72E-03
40	MLP 1-3-1	0.97	0.98	0.99	Exponential	Exponential	909.41	2052.58	2.80%	5.71%	2526.15	3.85E-03
40	MLP 1-2-1	0.97	0.96	0.99	Logistic	Logistic	1110.51	2279.94	3.35%	6.40%	2764.05	4.61E-03
40	MLP 1-8-1	0.97	0.96	0.99	Logistic	Exponential	1483.15	2363.70	4.57%	6.80%	2927.92	5.17E-03
40	MLP 1-3-1	0.96	0.95	0.99	Logistic	Exponential	1035.60	2590.23	3.60%	7.73%	3048.65	5.61E-03
60	MLP 1-4-1	0.97	0.97	0.99	Tanh	Logistic	1031.48	2377.94	3.10%	6.68%	2772.29	4.64E-03
60	MLP 1-7-1	0.97	0.97	0.99	Tanh	Logistic	777.04	2415.30	2.17%	6.63%	2763.52	4.61E-03
60	MLP 1-2-1	0.97	0.97	0.99	Exponential	Logistic	1109.93	2233.82	3.17%	6.12%	2715.75	4.45E-03
60	MLP 1-5-1	0.97	0.97	0.99	Tanh	Exponential	1090.27	2283.83	3.28%	6.40%	2721.35	4.47E-03
60	MLP 1-6-1	0.97	0.98	0.99	Exponential	Exponential	1071.39	1987.71	3.83%	5.94%	2590.07	4.05E-03
80	MLP 1-5-1	0.97	0.97	0.99	Tanh	Logistic	1040.10	2404.10	3.15%	6.78%	2796.23	4.72E-03
80	MLP 1-8-1	0.97	0.98	0.99	Exponential	Logistic	1023.90	2217.06	3.03%	6.15%	2661.45	4.27E-03
80	MLP 1-3-1	0.97	0.98	0.99	Exponential	Logistic	801.65	2239.47	2.48%	6.27%	2623.54	4.15E-03
80	MLP 1-7-1	0.97	0.96	0.99	Logistic	Logistic	978.70	2426.95	2.68%	6.63%	2851.96	4.91E-03
80	MLP 1-7-1	0.97	0.98	0.99	Exponential	Logistic	873.08	2237.49	2.54%	6.16%	2638.93	4.20E-03
100	MLP 1-8-1	0.97	0.98	0.99	Exponential	Logistic	1021.62	2260.14	2.99%	6.25%	2688.97	4.36E-03
100	MLP 1-5-1	0.97	0.97	0.99	Logistic	Logistic	1108.50	2402.26	3.24%	6.69%	2819.42	4.80E-03
100	MLP 1-4-1	0.97	0.98	0.99	Logistic	Exponential	909.58	2320.20	2.93%	6.65%	2707.86	4.43E-03
100	MLP 1-2-1	0.96	0.95	0.99	Tanh	Logistic	1114.15	2426.31	4.15%	7.43%	3005.79	5.45E-03
100	MLP 1-2-1	0.97	0.97	0.99	Tanh	Logistic	894.03	2347.67	2.83%	6.67%	2720.34	4.47E-03
200	MLP 1-6-1	0.96	0.96	0.99	Tanh	Logistic	644.88	2480.32	2.22%	7.15%	2814.19	4.78E-03
200	MLP 1-6-1	0.97	0.96	0.99	Tanh	Logistic	770.95	2330.09	2.51%	6.64%	2702.56	4.41E-03
200	MLP 1-3-1	0.97	0.97	0.99	Logistic	Logistic	970.77	2347.01	2.97%	6.61%	2750.08	4.56E-03
200	MLP 1-2-1	0.95	0.92	0.99	Logistic	Exponential	319.55	2657.15	0.32%	7.61%	3035.78	5.56E-03
200	MLP 1-4-1	0.97	0.97	0.99	Tanh	Logistic	1200.19	2356.37	3.63%	6.66%	2816.40	4.79E-03
						Minimal	**319.55**	**1987.71**	**0.32%**	**5.68%**	**2484.15**	**3.72E-03**

Table 2 – Summary of neural network learning for a random learning sample size of 80%, a test sample size of 10% and a validation sample size of 10% for Poland

Network number	Network name	Quality (learning)	Quality (testing)	Quality (validation)	Activation (hidden)	Activation (output)	Errors					
							ME	MAE	MPE	MAPE	SSE	Theil
20	MLP 1-5-1	0.96	0.99	1.00	Logistic	Linear	422.40	1830.32	0.90%	5.12%	2362.11	3.37E-03
20	MLP 1-5-1	0.96	0.99	1.00	Linear	Tanh	420.05	2152.64	0.39%	6.51%	2773.07	4.64E-03
20	MLP 1-3-1	0.96	0.99	1.00	Tanh	Logistic	702.37	1986.10	2.31%	5.57%	2455.03	3.64E-03
20	MLP 1-8-1	0.96	0.99	1.00	Linear	Tanh	326.74	2130.77	0.17%	6.45%	2734.11	4.51E-03
20	MLP 1-8-1	0.96	0.99	1.00	Logistic	Tanh	265.62	1759.88	0.80%	4.63%	2294.49	3.18E-03
40	MLP 1-5-1	0.96	0.99	1.00	Tanh	Exponential	1544.71	2539.20	6.27%	8.23%	3300.44	6.57E-03
40	MLP 1-5-1	0.96	0.99	1.00	Linear	Tanh	180.47	2355.28	0.72%	7.31%	2994.43	5.41E-03
40	MLP 1-2-1	0.96	0.99	1.00	Linear	Tanh	184.52	2325.08	1.67%	7.28%	2934.77	5.20E-03
40	MLP 1-6-1	0.96	0.98	1.00	Logistic	Logistic	725.12	2046.17	3.35%	6.03%	2699.40	4.40E-03
40	MLP 1-2-1	0.96	0.99	1.00	Logistic	Tanh	397.20	1761.51	1.05%	4.76%	2339.08	3.30E-03
60	MLP 1-2-1	0.95	0.98	1.00	Logistic	Exponential	46.12	2638.20	0.89%	7.75%	3021.27	5.51E-03
60	MLP 1-6-1	0.96	0.99	1.00	Linear	Tanh	381.15	2625.38	2.79%	8.41%	3359.27	6.81E-03
60	MLP 1-6-1	0.95	0.98	1.00	Logistic	Logistic	1436.54	2605.98	2.62%	6.58%	3107.29	5.83E-03
60	MLP 1-3-1	0.95	0.98	1.00	Tanh	Tanh	225.51	2181.93	1.10%	6.66%	2827.63	4.83E-03
60	MLP 1-6-1	0.95	0.99	1.00	Exponential	Logistic	231.31	2206.35	0.69%	5.98%	2657.70	4.26E-03
80	MLP 1-2-1	0.96	0.99	1.00	Logistic	Tanh	63.00	2068.87	0.35%	6.24%	2669.34	4.30E-03
80	MLP 1-3-1	0.96	0.99	1.00	Linear	Tanh	261.75	2325.06	0.42%	7.18%	2957.41	5.28E-03
80	MLP 1-2-1	0.96	0.98	1.00	Logistic	Linear	553.25	2205.41	2.23%	6.74%	2759.02	4.59E-03
80	MLP 1-2-1	0.95	0.98	1.00	Tanh	Logistic	81.51	2328.80	0.41%	6.55%	2719.89	4.46E-03
80	MLP 1-7-1	0.96	0.99	1.00	Linear	Tanh	159.97	2374.17	0.82%	7.38%	3018.42	5.50E-03
100	MLP 1-7-1	0.96	0.99	1.00	Linear	Tanh	573.15	2175.01	0.84%	6.54%	2792.33	4.71E-03
100	MLP 1-2-1	0.95	0.99	1.00	Tanh	Logistic	334.46	2310.29	1.71%	6.79%	2726.36	4.49E-03
100	MLP 1-5-1	0.96	0.99	1.00	Linear	Tanh	180.96	2441.21	1.90%	7.72%	3101.83	5.81E-03
100	MLP 1-2-1	0.96	0.99	1.00	Linear	Tanh	573.25	2174.78	0.84%	6.54%	2791.98	4.70E-03
100	MLP 1-4-1	0.96	0.99	1.00	Linear	Tanh	100.84	2331.11	0.91%	7.25%	2967.66	5.32E-03
200	MLP 1-8-1	0.96	0.99	1.00	Tanh	Tanh	380.18	2350.66	2.18%	7.47%	3034.06	5.56E-03
200	MLP 1-2-1	0.96	0.98	1.00	Tanh	Linear	265.66	2300.27	1.76%	7.12%	2877.84	5.00E-03
200	MLP 1-8-1	0.96	0.99	1.00	Tanh	Tanh	1932.17	2744.39	4.08%	6.86%	3486.86	7.34E-03
200	MLP 1-3-1	0.96	0.98	1.00	Logistic	Tanh	38.54	1969.49	0.44%	5.51%	2441.32	3.60E-03
200	MLP 1-4-1	0.95	0.98	1.00	Logistic	Logistic	704.60	2296.70	1.35%	6.17%	2731.82	4.50E-03
						Minimal	**38.54**	**1759.88**	**0.17%**	**4.63%**	**2294.49**	**3.18E-03**

Table 3 – Summary of neural network learning for a random learning sample size of 70%, a test sample size of 15% and a validation sample size of 15% for Slovenia

Network number	Network name	Quality (learning)	Quality (testing)	Quality (validation)	Activation (hidden)	Activation (output)	Errors					
							ME	MAE	MPE	MAPE	SSE	Theil
20	MLP 1-4-1	0.75	1.00	1.00	Linear	Linear	30.71	193.48	0.91%	3.56%	372.79	3.45E-03
20	MLP 1-7-1	0.75	1.00	1.00	Logistics	Logistics	133.34	260.48	2.84%	4.80%	457.86	3.89E-03
20	MLP 1-8-1	0.74	1.00	1.00	Linear	Tanh	164.66	265.45	3.36%	4.91%	464.21	1.43E-01
20	MLP 1-5-1	0.77	1.00	1.00	Exponential	Linear	3.69	185.43	0.37%	3.38%	359.54	1.49E-01
20	MLP 1-3-1	0.72	1.00	1.00	Exponential	Exponential	6.30	198.22	0.60%	3.61%	374.35	4.39E-03
40	MLP 1-5-1	0.73	1.00	1.00	Tanh	Tanh	6.47	191.92	0.51%	3.52%	369.30	3.45E-03
40	MLP 1-4-1	0.71	1.00	1.00	Logistics	Tanh	56.36	192.69	1.36%	3.56%	379.19	3.89E-03
40	MLP 1-2-1	0.76	1.00	1.00	Tanh	Linear	38.47	283.70	1.16%	5.12%	436.69	1.43E-01
40	MLP 1-4-1	0.75	1.00	1.00	Linear	Linear	28.74	189.75	0.89%	3.49%	370.73	1.49E-01
40	MLP 1-5-1	0.74	1.00	1.00	Tanh	Tanh	37.63	175.76	1.06%	3.39%	368.68	4.39E-03
60	MLP 1-4-1	0.77	1.00	1.00	Tanh	Exponential	8.34	187.24	0.29%	3.40%	358.98	3.45E-03
60	MLP 1-4-1	0.77	1.00	1.00	Tanh	Logistics	14.49	187.31	0.67%	3.42%	359.24	3.89E-03
60	MLP 1-7-1	0.74	1.00	1.00	Logistics	Tanh	39.18	291.45	1.17%	5.26%	447.96	1.43E-01
60	MLP 1-3-1	0.77	1.00	1.00	Exponential	Tanh	12.82	196.63	0.21%	3.56%	357.07	1.49E-01
60	MLP 1-7-1	0.77	1.00	1.00	Linear	Exponential	8.22	186.00	0.29%	3.38%	359.40	4.39E-03
80	MLP 1-8-1	0.72	1.00	1.00	Tanh	Tanh	27.43	193.68	0.86%	3.57%	375.90	3.45E-03
80	MLP 1-4-1	0.77	1.00	1.00	Logistics	Logistics	16.09	200.87	0.15%	3.62%	355.34	3.89E-03
80	MLP 1-4-1	0.77	1.00	1.00	Exponential	Tanh	30.00	273.63	1.01%	4.94%	420.37	1.43E-01
80	MLP 1-6-1	0.77	1.00	1.00	Tanh	Logistics	14.31	189.16	0.19%	3.43%	358.57	1.49E-01
80	MLP 1-7-1	0.74	1.00	1.00	Exponential	Exponential	57.82	217.51	1.53%	4.00%	408.16	4.39E-03
100	MLP 1-4-1	0.74	1.00	1.00	Tanh	Linear	2.56	191.36	0.42%	3.48%	367.28	3.45E-03
100	MLP 1-6-1	0.77	1.00	1.00	Tanh	Exponential	9.40	187.94	0.27%	3.41%	358.87	3.89E-03
100	MLP 1-3-1	0.76	1.00	1.00	Exponential	Linear	95.83	242.44	2.16%	4.47%	425.61	1.43E-01
100	MLP 1-8-1	0.74	1.00	1.00	Exponential	Exponential	70.51	222.33	1.75%	4.09%	416.18	1.49E-01
100	MLP 1-3-1	0.77	1.00	1.00	Tanh	Linear	41.13	189.96	1.14%	3.49%	362.66	4.39E-03
200	MLP 1-8-1	0.75	1.00	1.00	Linear	Exponential	85.16	219.40	1.99%	4.06%	415.90	3.45E-03
200	MLP 1-6-1	0.74	1.00	1.00	Exponential	Exponential	55.75	211.18	1.48%	3.89%	400.48	3.89E-03
200	MLP 1-7-1	0.76	1.00	1.00	Exponential	Logistics	25.94	300.56	0.93%	5.40%	450.20	1.43E-01
200	MLP 1-2-1	0.76	1.00	1.00	Exponential	Linear	32.63	189.13	1.02%	3.49%	366.05	1.49E-01
200	MLP 1-5-1	0.74	1.00	1.00	Exponential	Exponential	14.96	197.62	0.21%	3.58%	368.45	4.39E-03
						Minimal	**2.56**	**175.76**	**0.15%**	**3.38%**	**355.34**	**3.45E-03**

Table 4 – Summary of neural network learning for a random learning sample size of 80%, a test sample size of 10% and a validation sample size of 10% for Slovenia

Network number	Network name	Quality (learning)	Quality (testing)	Quality (validation)	Activation (hidden)	Activation (output)	Errors					Thail
							ME	MAE	MPE	MAPE	SSE	
20	MLP 1-7-1	0.78	0.00	0.00	Logistics	Tanh	10.52	188.64	0.25%	3.42%	349.31	3.45E-03
20	MLP 1-7-1	0.75	0.00	0.00	Linear	Linear	24.97	193.64	0.82%	3.56%	371.72	3.89E-03
20	MLP 1-2-1	0.77	0.00	0.00	Logistics	Logistics	14.25	274.47	0.72%	4.93%	409.26	1.43E-01
20	MLP 1-8-1	0.72	0.00	0.00	Linear	Tanh	206.55	242.07	3.95%	4.50%	442.47	1.49E-01
20	MLP 1-8-1	0.77	0.00	0.00	Linear	Exponential	12.84	187.08	0.64%	3.41%	359.00	4.39E-03
40	MLP 1-5-1	0.78	0.00	0.00	Logistics	Logistics	11.65	206.39	0.23%	3.72%	352.99	3.45E-03
40	MLP 1-7-1	0.77	0.00	0.00	Exponential	Exponential	3.42	192.99	0.37%	3.50%	355.22	3.89E-03
40	MLP 1-3-1	0.82	0.00	0.00	Logistics	Logistics	23.22	249.17	0.76%	4.43%	367.46	1.43E-01
40	MLP 1-8-1	0.77	0.00	0.00	Logistics	Logistics	4.78	193.91	0.52%	3.53%	356.38	1.49E-01
40	MLP 1-6-1	0.77	0.00	0.00	Exponential	Logistics	43.43	198.40	1.20%	3.65%	365.75	4.39E-03
60	MLP 1-4-1	0.77	0.00	0.00	Logistics	Linear	3.45	197.39	0.37%	3.57%	355.30	3.45E-03
60	MLP 1-4-1	0.75	0.00	0.00	Exponential	Exponential	14.41	200.11	0.63%	3.66%	371.35	3.89E-03
60	MLP 1-6-1	0.77	0.00	0.00	Logistics	Exponential	8.63	274.84	0.62%	4.92%	405.66	1.43E-01
60	MLP 1-6-1	0.75	0.00	0.00	Exponential	Exponential	15.80	201.76	0.65%	3.69%	372.48	1.49E-01
60	MLP 1-2-1	0.76	0.00	0.00	Linear	Exponential	285.98	309.26	5.36%	5.71%	514.53	4.39E-03
80	MLP 1-5-1	0.77	0.00	0.00	Logistics	Exponential	6.35	195.89	0.32%	3.55%	356.70	3.45E-03
80	MLP 1-7-1	0.76	0.00	0.00	Tanh	Logistics	9.84	190.05	0.57%	3.46%	361.92	3.89E-03
80	MLP 1-2-1	0.77	0.00	0.00	Exponential	Tanh	14.55	274.13	0.73%	4.92%	409.86	1.43E-01
80	MLP 1-6-1	0.77	0.00	0.00	Exponential	Logistics	5.75	197.31	0.33%	3.57%	356.63	1.49E-01
80	MLP 1-2-1	0.75	0.00	0.00	Linear	Logistics	4.83	198.72	0.47%	3.62%	366.91	4.39E-03
100	MLP 1-3-1	0.75	0.00	0.00	Linear	Logistics	5.52	203.46	0.29%	3.69%	367.79	3.45E-03
100	MLP 1-3-1	0.75	0.00	0.00	Linear	Linear	30.47	193.10	0.91%	3.55%	372.55	3.89E-03
100	MLP 1-3-1	0.73	0.00	0.00	Linear	Tanh	224.97	298.81	4.35%	5.54%	506.05	1.43E-01
100	MLP 1-7-1	0.77	0.00	0.00	Exponential	Logistics	4.41	196.64	0.35%	3.56%	356.62	1.49E-01
100	MLP 1-2-1	0.72	0.00	0.00	Linear	Tanh	60.29	202.17	1.41%	3.73%	384.30	4.39E-03
200	MLP 1-5-1	0.76	0.00	0.00	Logistics	Logistics	1.55	199.88	0.36%	3.63%	365.81	3.45E-03
200	MLP 1-2-1	0.77	0.00	0.00	Logistics	Tanh	8.82	197.37	0.28%	3.57%	355.53	3.89E-03
200	MLP 1-8-1	0.76	0.00	0.00	Tanh	Logistics	44.49	285.58	1.24%	5.15%	432.62	1.43E-01
200	MLP 1-2-1	0.66	0.00	0.00	Logistics	Linear	105.71	226.65	2.17%	4.18%	409.52	1.49E-01
200	MLP 1-5-1	0.75	0.00	0.00	Linear	Logistics	4.76	198.82	0.47%	3.62%	366.94	4.39E-03
						Minimal	**1.55**	**187.08**	**0.23%**	**3.41%**	**349.31**	**3.45E-03**

Despite a possible minor uptick, the research predicts that Poland's traffic accident rate will level out in the upcoming years. The sample sizes selected for model training, validation and testing are among the variables that affect how accurate these projections are. In general, prediction accuracy may be raised by increasing the percentage of data allotted for model training. For example, increasing the size of the training set to 80% (80-10-10 split) reduced the average percentage error from 5.68% (using a 70-15-15 split across training, testing and validation sets) to 4.63%. In addition, the increasing number of cars on Polish roads and the recent COVID-19 epidemic have an impact on the model's forecasts (*Figure 3*).
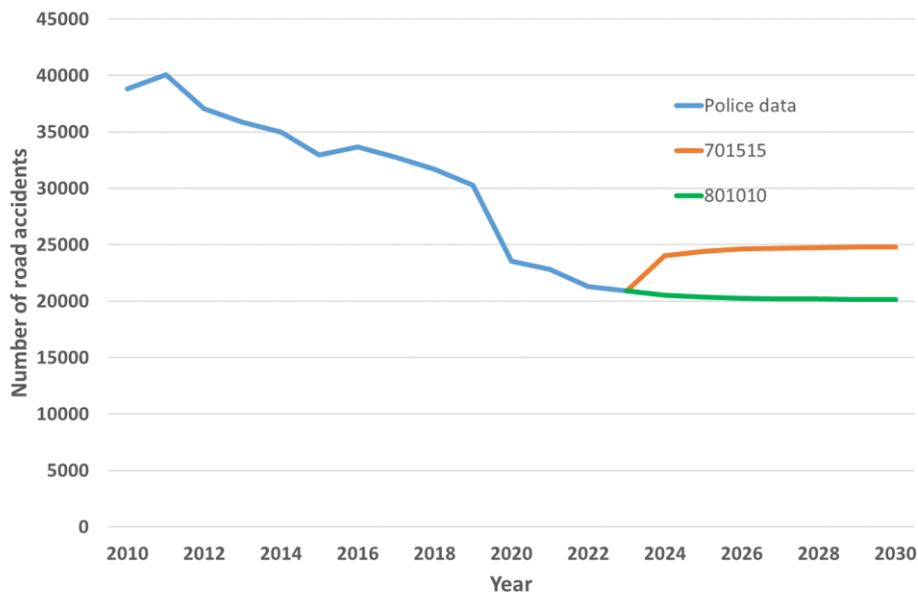
Figure 3 – Projected number of road accidents in Poland between 2024 and 2030

However, Slovenia showed the opposite pattern, with traffic accidents steadily increasing until 2020, when a steep decline was noted as a result of the COVID-19 lockdown. But once the market reopened, there was a sharp rise in traffic accidents. When the training set size was extended to 80% (80-10-10 split), the average percentage error for forecasting the traffic accident trend for Slovenia increased somewhat from 2.14% with a 70-15-15 split between training, testing and validation sets to 4.84%. Although there are several factors influencing road accidents and the COVID epidemic has a major impact on traffic reduction, the forecasting approach is effective at predicting future traffic accident trends with the fewest possible mistakes (*Figure 4*).
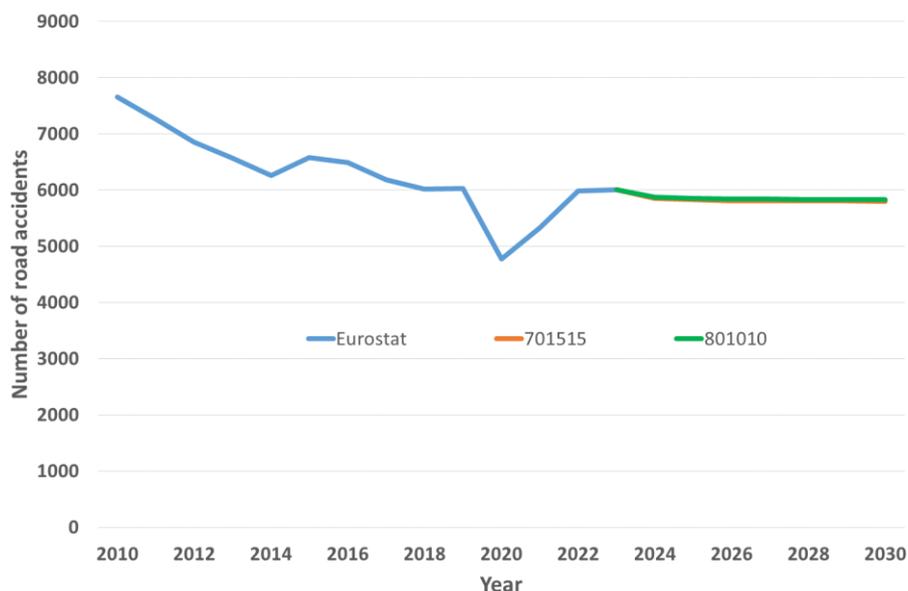

Figure 4 – Projected number of road accidents in Slovenia between 2024 and 2030

4. CONCLUSION

This study predicted the frequency of traffic accidents in Poland and Slovenia using neural network models in the Statistica environment. In order to reduce the mean absolute error and mean absolute percentage error, the algorithm adjusted the network weights.

Based on factors such as the recent pandemic and the increasing number of cars on the road, the results indicate that the number of traffic accidents will stabilise in the coming years. By examining the calculated prediction errors, the accuracy of the predictions was evaluated.

In light of these predictions, it is necessary to take proactive steps to further minimise the number of traffic accidents. Recent steps have been taken in this direction, such as the introduction of higher fines for traffic offenses in Poland (which will go into effect on 1 January 2022).

The COVID-19 epidemic had a major impact on traffic patterns, which to some extent distorted the survey data collected.

Future studies will examine the use of different statistical approaches and the inclusion of other variables affecting accident frequency, such as traffic volume, weather, driver age and the use of exponential smoothing techniques.

REFERENCES

[1]   Abdullah E, Emam A. Traffic accidents analyzer using big data. In: *Proceedings of the 2015 International Conference on Computational Science and Computational Intelligence (CSCI 2015).* Institute of Electrical and Electronics Engineers (IEEE); 2016. p. 392-397. DOI: 10.1109/CSCI.2015.187

[2]   Al-Madani H. Global road fatality trends'estimations based on country-wise microlevel data, *Accident Analysis & Prevention*. 2018;111:297–310, DOI: 10.1016/j.aap.2017.11.035

[3]   Arteaga C, Paz A, Park J. Injury severity on traffic crashes: A text mining with an interpretable machine-learning approach. *Safety Science*. 2020;132:104988. DOI: 10.1016/j.ssci.2020.104988

[4]   Bąk I, Cheba K, Szczecińska B. The statistical analysis of road traffic in cities of Poland. *Transportation Research Procedia*, 2019;39:14-23, DOI: 10.1016/j.trpro.2019.06.003

[5]   Biswas AA, Mia J, Majumder A. Forecasting the Number of Road Accidents and Casualties using Random Forest Regression in the Context of Bangladesh. In *Proceedings of the 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT),* Kanpur, India, 6-8 July 2019.

[6]   Bloomfield P. An exponential model in the spectrum of a scalar time series. *Biometrika* 1973;60:217–226 https://www.jstor.org/stable/2334533

[7]   Central Statistical Office. 2022. www.gus.pl [Accessed 12th May 2022].

[8]   Chand A, Jayesh S, Bhasi AB. Road traffic accidents: An overview of data sources, analysis techniques and contributing factors, *Materials Today: Proceedings*. 2021:47:5135-5141. DOI: 10.1016/j.matpr.2021.05.415

[9]   Chen C. *Analysis and forecast of traffic accident big data*. ITM Web Conf., 2017:12:04029. DOI: 10.1051/itmconf/20171204029

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

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

[12] Data mining techniques StatSoft. 2022. https://www.statsoft.pl/textbook/stathome_stat.html?https%3A%2F%2Fwww.statsoft.pl%2Ftextbook%2Fstdatmin.html [Accessed 12th May 2022].

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

[14] Dudek G. *Forecasting time series with multiple seasonal cycles using neural networks with local learning.* In: Rutkowski L, et al. (eds) Artificial Intelligence and Soft Computing. ICAISC 2013. Lecture Notes in Computer Science,7894. Springer, Berlin, Heidelberg. DOI: 10.1007/978-3-642-38658-9_5

[15] Dutta B, Barman MP, Patowary AN. Application of Arima model for forecasting road accident deaths in India. *International Journal of Agricultural and Statistical Sciences*. 2020;16:2:607-615.

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

[17] Fiszeder P. *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 [Accessed 12th May 2022].

[19] Gorzelanczyk P, Pyszewska D, Kalina T, Jurkovic M. Analysis of road traffic safety in the Pila poviat. *Scientific Journal of Silesian University of Technology. Series Transport.* 2020;107:33-52. DOI: 10.20858/sjsutst.2020.107.3

[20] Helgason A, Fractional integration methods and short time series: Evidence from asimulation study, *Political Analysis.* 2016;24:1:59–68. http://www.jstor.org/stable/24573204

[21] Statistic Road Accident 2022 https://statystyka.policja.pl/ [Accessed 12th May 2022].

[22] Eurostat 2022 https://ec.europa.eu/eurostat/data/database [Accessed 12th May 2022].

[23] Lake BM, Ullman TD, Tenenbaum JB, Gershman SJ. Building machines that learn and think like people. *Behaviour Brain Sciences* 2017;40:e253.

[24] Becoming Human. How Netflix uses AI, data science, and machine learning – from a product perspective. 2022. www.becominghuman.ai [Accessed 12th May 2022].

[25] Forbes. The amazing ways eBay is using Artificial Intelligence to boost business success. 2022. www.forbes.com [Accessed 12th May 2022].

[26] Wu Y, et al. Google's neural machine translation system: Bridging the gap between human and machine translation. *ArXiv*. 2022;1609.08144. www.arxiv.org [Accessed 12th May 2022].

[27] Oronowicz-Jaśkowiak W. The application of neural networks in the work of forensic experts in child abuse cases. *Advances in Psychiatry and Neurology.* 2019;28(4):273-282. DOI: 10.5114/ppn.2019.92489