https://doi.org/10.55343/CogSust.102

Forecasting the number of road accidents caused by pedestrians in Poland using neural networks

Piotr Gorzelańczyk
https://orcid.org/0000-0001-9662-400X
Department of Transport. Stanislaw Staszic State University of Applied Sciences in Piła, Podchorazych 10 Street, 64-920 Piła, Poland. E-mail: piotr.gorzelanczyk@ans.pila.pl;

Abstract

Every year, fewer traffic accidents occur in Poland and throughout the world. Pandemics have recently impacted this number, but it is still relatively high. All efforts should be made to lower this figure. The article's main goal is to project the number of pedestrian-related traffic accidents in Poland based on yearly statistics. from 2001. A projection for the years 2024–2030 was created using police data. Various neural network models were employed to predict the number of incidents. The findings indicate that a stabilisation in traffic accidents is yet to be expected. One way to look at this is as a result of both Poland's population reduction and the growing number of cars on the road. The number of random samples (training, test, and validation) selected has little effect on the outcomes (Road safety statistics in the EU, 2024, Poland Population, 2024, Poland Number of Registered, 2024).

Keywords

road accident, pandemic, forecasting, neural networks, pedestrian

1. Introduction

Road accidents occurred when at least one of the parties engaged was a moving vehicle that resulted in property damage or personal harm in a location meant for public transportation or widely utilised. The WHO estimates that 1.3 million individuals lose their lives in car crashes each year. Worldwide, the average country experiences a 3% decline in GDP due to traffic accidents. For children and young adults between the ages of five and twenty-nine, traffic accidents constitute the main cause of mortality (WHO, 2018). By 2030, the UN General Assembly wants a 50% decrease in traffic accident fatalities and injuries.

When judging how serious a traffic incident is, one factor to consider is its extent (Baranyai and Sipos, 2022). For the responsible authorities to develop road safety legislation aimed at averting accidents and reducing injuries, fatalities, and property damage, the severity of incidents must be estimated (Tambouratzis et al., 2014; Zhu et al., 2019; Mekonnen et al., 2022).

Before implementing countermeasures to prevent or reduce the number of serious road accidents, it is imperative to identify the critical elements that influence accident severity (Arteaga et al., 2020). A multi-node Deep Neutral Network (DNN) architecture is presented by Yang et al. (2022) to forecast varying degrees of injury, death, and property damage. It makes it possible to accurately and thoroughly assess the severity of traffic accidents.

The accident figures come from several sources. Usually, government agents use the appropriate government entities to collect and evaluate them. Numerous sources are used to collect data, such as police reports, insurance company databases, and hospital records. As a result, the transportation industry is doing an extensive data analysis on traffic accidents (Gorzelanczyk et al., 2020).

The most significant information source for analysing and forecasting traffic events is intelligent transportation systems. GPS equipment mounted on moving cars may be used to analyse this data (Chen, 2017). Roadside microwave vehicle detection systems can continually capture information about moving automobiles, such as vehicle type, speed, and traffic volume (Khaliq et al., 2019). A license plate recognition system may likewise gather large amounts of traffic data over a predetermined time (Rajput et al., 2015). Social media is another possible source of information on traffic and accidents, albeit the reporters' lack of expertise may make their reporting less accurate (Zheng et al., 2018)

https://doi.org/10.55343/CogSust.102

Working with various data sources presents challenges that must be addressed before accident data can be considered useful. Accurate analytical findings can be achieved by merging heterogeneous traffic accident data and combining multiple data sources (Abdullah & Emam, 2015).

In order to determine the severity of the issue and establish a connection between traffic participants and accidents, Vilaça et al. (2017) conducted a statistical analysis. The study suggests enacting additional traffic safety measures and raising the bar for traffic law requirements.

A statistical examination of traffic safety in a chosen Polish area was conducted by Bąk et al. (2019) using the number of traffic accidents as a proxy for the amount of research on accident causes. Multivariate statistical analysis was employed in the study to examine the safety variables of accident causes.

The sort of traffic issue being handled determines the source of accident data to use for analysis. Combining statistical models with additional data from real driving or other information gleaned from intelligent traffic systems enhances accident prediction and accident eradication accuracy (Chand et al., 2021).

The literature offers several techniques for predicting the probability of accidents. Time series approaches are the most popular techniques for forecasting accident frequency (Helgason, 2016; Lavrenz et al., 2018). However, they have the drawback of being unable to evaluate the forecast's accuracy based on previous predictions and frequently having a residual autocorrelation component (Forecasting based on time series, n. d.). While Sunny et al. (2018) employed the Holt-Winters exponential smoothing approach, Procházka et al. (2017) used a multi-seasonality model for their forecasts. One of the model's drawbacks is that exogenous variables cannot be included (Al-Madani, 2018).

The frequency of traffic accidents has been predicted using curve-fitting regression models of Al-Madani (2018) and Monedero et al. (2021) for analysing the number of fatalities, as well as the vector autoregressive model, which has the disadvantage of requiring many observations of variables to estimate their parameters (Wójcik, 2014) accurately. In turn, these need only a few basic linear connections (Piłatowska, 2012) and an order of autoregression (Mamczur, 2020), supposing the series is already stable.

Random Forest regression was used by Biswas et al. (2019) to forecast the frequency of traffic accidents. The data comprise groups with linked features that are similarly relevant to the original data, the approach and peak prediction are unstable, and smaller groups are preferred over bigger ones in this case. For the given forecasting problem, Chudy-Laskowska and Pisula (2015) employed an autoregressive quadratic trend model, a univariate periodic trend model, and an exponential smoothing model. A moving average model may also be utilised to foresee the problem. However, this approach has drawbacks, such as poor forecast accuracy, data loss within a sequence, and an inability to consider trends and seasonal variations (Kashpruk, 2020).

To ensure that the process remains stable, Prochazka and Camaj (2017) employed the GARMA approach, which confines the parameter space. Forecasting frequently uses the ARMA model for stationary systems (Dutta et al., 2020; Karlaftis et al., 2009); it uses the ARIMA or SARIMA model for non-stationary processes. These approaches have the benefit of giving the models under examination considerable flexibility. However, they also have the drawback of requiring more research competence from the researcher than, say, regression analysis (Łobejko, 2015). The linearity of the ARIMA model is another drawback (Szmuksta-Zawadzka & Zawadzki, 2009).

An ANOVA test was employed in Chudy-Laskowska and Pisula's work (2015) to forecast the frequency of traffic accidents. This approach's drawback is that it makes several extra assumptions. The most important among them is the sphericity premise, which might result in incorrect findings. Neural network techniques are also used to forecast the frequency of vehicle accidents. One of the drawbacks of artificial neural networks (ANNs) is that they require prior expertise in this area (Wrobel, 2017), as the final result depends on the network's initial conditions. Additionally, because ANNs are often called "black boxes", where input data is entered and the model outputs results without knowing the analysis, it is impossible to interpret results conventionally (Data mining techniques, n. d.).

Kumar et al. (2019) used the Hadoop model as a state-of-the-art prediction technique. This strategy's drawback is its inability to handle tiny data sets (Top Advantages and Disadvantages of Hadoop 3, n. d.). Karlaftis and Vlahogianni (2009) forecasted using the Garch model. This strategy's intricate model and form provide a problem (Perczak & Fiszeder, 2014). Nevertheless, McIlroy and his team's ADF test has the drawback of having insufficient power to detect the autocorrelation of the random component (McIlroy et al., 2019).

https://doi.org/10.55343/CogSust.102

Another significant factor is the number of accidents per 10,000 people (NRA). In Poland, where there are 37.6 million people, there were 20,936 traffic accidents in 2023. By entering the data into equation 1, we can determine that in Poland in 2023, there were 5.56 traffic accidents per 1,000 people that a pedestrian caused.

$$NRA = \frac{NR}{NI} * 1000 \qquad (1)$$

where:

NR – number of road accidents

NI – number of inhabitants

An accident involving a pedestrian is a traffic incident in which a pedestrian is hit by a vehicle. The most common cause of this type of accident is a vehicle driver's failure to prioritise a pedestrian.

Based on the data above, the author projected the number of pedestrian-related accidents on Polish roadways. The amount of accidents was anticipated using neural networks.

2. Materials and methods

Pedestrians are involved in a significant number of traffic accidents each year. The epidemic has decreased the number of traffic accidents in recent years, which has impacted the predicted value. Pedestrians remain the primary cause of many traffic accidents, even during the pandemic. Because of this, every effort should be made to lower this number and show how it will change over the next several years (Figure 1):

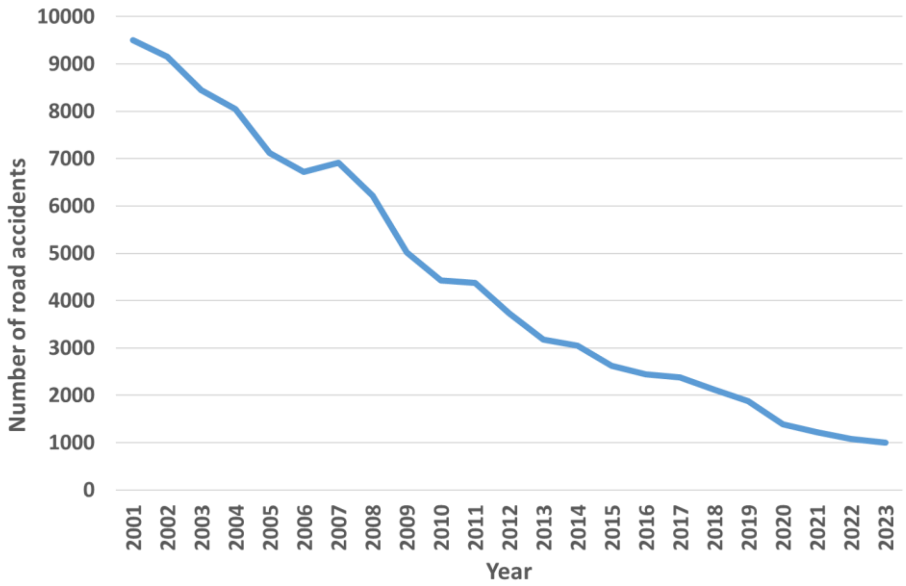

Figure 1. Number of road accidents in Poland caused by pedestrians in years 2001-2023 (Road Accident Statistics, 2024)

The number of traffic accidents in Poland was predicted using a subset of neural network models. One benefit of this approach is that it mimics how the human brain functions. A neural network comprises nodes with inputs, weights, variances, and outputs. The Statistica software was used to choose the ideal weights throughout the investigation. The model and parameters used will determine the outcome of the forecast made using this approach.

One way to think of a neural network is as a mathematical construct that functions by drawing on the nervous system. They usually consist of several layers that together form the network's architecture. Thanks to a procedure known as training, the first layer includes data regarding text, pictures, numbers, and sounds. The network may use hundreds of inputs in this procedure before drawing specific judgments. Artificial neurons, which are mathematical functions that mimic the activity of organic neurons, are the fundamental building blocks of neural networks. Artificial neurons have several inputs but only give one output value, like organic neurons. When it comes to biological neurons, this is comparable to dendrites. Artificial intelligence development is centred around neural networks. Rather, this field of study aims to develop models that use intelligent behaviour, such as

https://doi.org/10.55343/CogSust.102

knowledge generalisation (Lake et al., 2017). Forecasting using neural networks extends to the frequency of traffic incidents (Oronowicz-Jaśkowiak, 2019).

The Statistica software and its built-in artificial neural network modules corrected the ideal weights throughout testing. A multilayer perceptron (MLP) neural network, including layers of hidden neurons, was employed for prediction. In the cases under analysis, the number of neurons in the intermediate layer varied between two and eight. A single neuron in the output layer represented the time series output values of the number of traffic accidents. The model and model parameters used will determine the predictive outcomes of the approaches presented. The measure of predictive brilliance was derived from the following prediction errors, which were found using equations (2–7).

- *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

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

- *SSE* – mean square error

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

- M^2 – Theil's 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 of forecast horizon,

Y – the observed value of road accidents,

Y_p – the forecasted value of road accidents.

Neural network models with the lowest mean and absolute percentage errors were utilised to forecast the frequency of traffic incidents caused by pedestrians.

3. Results

Data from the Polish Police from 2001 to 2023 were used to anticipate the number of pedestrian-related traffic accidents in Poland each year. The study was carried out using the *Statistica* software, assuming two random sample sizes:

1. training 70%, test 15% and validation 15%.
2. training 80%, test 10% and validation 10%,

with the following number of learning networks: 20, 40, 60, 80, 100, and 200, the MP error value was minimal (Tables 1–2). Consequently, the row for which the MAPE error value is the smallest is highlighted in every table.

https://doi.org/10.55343/CogSust.102

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

Network number	Network name	Quality (training)	Quality (testing)	Quality (validation)	Learning algorithm	Error function	Activation (hidden)	Activation (output)	Errors					
									ME	MAE	MPE	MAPE	SSE	Theil
20	MLP 1-7-1	0.986150	0.999572	0.999618	BFGS 58	SOS	Logistics	Linear	10.75	118.67	0.64%	5.13%	158.68	0.0035
20	MLP 1-3-1	0.985445	0.999616	0.999819	BFGS 30	SOS	Tanh	Tanh	0.91	108.71	0.69%	4.34%	158.32	0.0039
20	MLP 1-8-1	0.986154	0.999553	0.999526	BFGS 38	SOS	Exponential	Linear	13.69	123.14	1.03%	5.53%	160.56	0.1426
20	MLP 1-7-1	0.985622	0.999519	0.999496	BFGS 30	SOS	Logistics	Tanh	3.40	112.19	0.37%	4.70%	156.76	0.1489
20	MLP 1-2-1	0.986174	0.999562	0.999561	BFGS 115	SOS	Tanh	Linear	12.08	122.62	0.89%	5.41%	159.89	0.0044
40	MLP 1-8-1	0.986163	0.999162	0.999959	BFGS 46	SOS	Exponential	Logistics	11.43	105.37	0.80%	4.24%	145.17	0.0035
40	MLP 1-8-1	0.985171	0.999716	0.999998	BFGS 31	SOS	Exponential	Tanh	1.33	125.22	2.12%	6.00%	175.96	0.0039
40	MLP 1-4-1	0.986051	0.999316	0.999991	BFGS 39	SOS	Tanh	Logistics	2.04	93.60	0.34%	3.70%	147.77	0.1426
40	MLP 1-7-1	0.985798	0.999664	0.999999	BFGS 10	SOS	Exponential	Linear	10.26	111.76	0.99%	4.81%	166.33	0.1489
40	MLP 1-3-1	0.986067	0.999361	0.999920	BFGS 98	SOS	Tanh	Exponential	22.27	148.97	1.59%	6.42%	181.60	0.0044
60	MLP 1-6-1	0.986325	0.999100	0.999991	BFGS 66	SOS	Logistics	Logistics	2.37	97.62	0.20%	3.84%	144.97	0.0035
60	MLP 1-2-1	0.983635	0.999799	0.999990	BFGS 65	SOS	Logistics	Tanh	39.87	135.60	0.26%	6.24%	191.27	0.0039
60	MLP 1-6-1	0.986157	0.999497	1.000000	BFGS 39	SOS	Exponential	Exponential	32.95	129.90	3.14%	6.19%	167.26	0.1426
60	MLP 1-6-1	0.986250	0.998836	0.999990	BFGS 139	SOS	Exponential	Logistics	1.77	97.97	0.24%	3.94%	151.12	0.1489
60	MLP 1-8-1	0.985932	0.999524	0.999994	BFGS 27	SOS	Logistics	Exponential	89.93	163.09	7.00%	9.09%	206.06	0.0044
80	MLP 1-6-1	0.986358	0.999241	0.999987	BFGS 125	SOS	Exponential	Logistics	0.60	105.92	0.03%	4.16%	143.98	0.0035
80	MLP 1-4-1	0.986446	0.999461	0.999990	BFGS 93	SOS	Tanh	Logistics	3.47	100.47	0.14%	3.87%	141.03	0.0039
80	MLP 1-5-1	0.985687	0.999678	0.999982	BFGS 10	SOS	Exponential	Linear	8.30	123.63	2.35%	5.89%	173.19	0.1426
80	MLP 1-8-1	0.986897	0.999011	0.999997	BFGS 128	SOS	Tanh	Logistics	16.80	98.26	0.92%	3.95%	149.29	0.1489
80	MLP 1-3-1	0.984466	0.999551	0.999990	BFGS 38	SOS	Logistics	Logistics	36.41	131.87	4.22%	6.88%	185.54	0.0044
100	MLP 1-6-1	0.986010	0.998618	0.999978	BFGS 27	SOS	Tanh	Exponential	10.05	97.84	0.11%	3.98%	151.64	0.0035
100	MLP 1-7-1	0.985505	0.999306	1.000000	BFGS 14	SOS	Exponential	Logistics	8.95	131.35	2.04%	6.20%	167.54	0.0039
100	MLP 1-3-1	0.985645	0.999704	0.999994	BFGS 70	SOS	Exponential	Tanh	4.06	124.70	2.10%	5.88%	172.86	0.1426
100	MLP 1-7-1	0.984607	0.998820	0.999976	BFGS 27	SOS	Tanh	Logistics	51.42	143.59	4.92%	7.28%	194.61	0.1489
100	MLP 1-8-1	0.985894	0.999662	0.999997	BFGS 5	SOS	Exponential	Linear	16.41	120.91	2.13%	5.36%	164.46	0.0044
200	MLP 1-3-1	0.985868	0.999317	0.999990	BFGS 23	SOS	Tanh	Exponential	19.32	105.97	2.26%	4.70%	156.12	0.0035

https://doi.org/10.55343/CogSust.102

Network number	Network name	Quality (training)	Quality (testing)	Quality (validation)	Learning algorithm	Error function	Activation (hidden)	Activation (output)	Errors					
									ME	MAE	MPE	MAPE	SSE	Theil
200	MLP 1-3-1	0.985318	0.999582	0.999994	BFGS 45	SOS	Logistics	Logistics	67.28	148.61	5.33%	7.75%	185.06	0.0039
200	MLP 1-5-1	0.986394	0.999298	0.999993	BFGS 75	SOS	Logistics	Logistics	0.35	98.57	0.04%	3.82%	141.47	0.1426
200	MLP 1-6-1	0.986375	0.999501	0.999995	BFGS 147	SOS	Tanh	Logistics	1.15	99.08	0.02%	3.78%	140.57	0.1489
200	MLP 1-2-1	0.986022	0.999476	0.999996	BFGS 38	SOS	Exponential	Exponential	1.50	109.98	1.12%	4.66%	155.36	0.0044
								Minimal	**0.35**	**93.60**	**0.02%**	**3.70%**	**140.57**	**0.0035**

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

Network number	Network name	Quality (training)	Quality (testing)	Quality (validation)	Learning algorithm	Error function	Activation (hidden)	Activation (output)	Errors					
									ME	MAE	MPE	MAPE	SSE	Theil
20	MLP 1-2-1	0.985248	1.000000	1.000000	BFGS 21	SOS	Logistics	Tanh	12.66	118.11	1.33%	5.43%	175.00	0.0035
20	MLP 1-8-1	0.982742	1.000000	1.000000	BFGS 13	SOS	Tanh	Logistics	47.05	199.46	8.00%	12.47%	289.52	0.0039
20	MLP 1-5-1	0.983835	1.000000	1.000000	BFGS 27	SOS	Tanh	Tanh	16.06	140.85	3.04%	7.02%	185.14	0.1426
20	MLP 1-8-1	0.987041	1.000000	1.000000	BFGS 144	SOS	Exponential	Logistics	3.67	96.72	0.07%	3.75%	140.34	0.1489
20	MLP 1-6-1	0.986373	1.000000	1.000000	BFGS 15	SOS	Logistics	Linear	0.61	119.18	1.64%	5.39%	169.11	0.0044
40	MLP 1-6-1	0.984883	1.000000	1.000000	BFGS 41	SOS	Logistics	Tanh	16.06	123.82	2.45%	5.71%	172.02	0.0035
40	MLP 1-7-1	0.985665	1.000000	1.000000	BFGS 28	SOS	Tanh	Tanh	9.93	121.73	2.25%	5.66%	170.86	0.0039
40	MLP 1-3-1	0.954971	1.000000	1.000000	BFGS 6	SOS	Linear	Exponential	188.25	428.49	20.90%	26.88%	575.59	0.1426
40	MLP 1-2-1	0.986078	1.000000	1.000000	BFGS 62	SOS	Exponential	Tanh	14.73	115.27	1.00%	5.12%	170.69	0.1489
40	MLP 1-4-1	0.985140	1.000000	1.000000	BFGS 25	SOS	Logistics	Tanh	18.13	111.72	0.15%	4.97%	174.49	0.0044
60	MLP 1-4-1	0.956187	1.000000	1.000000	BFGS 6	SOS	Linear	Exponential	219.06	435.85	22.40%	27.74%	586.77	0.0035
60	MLP 1-7-1	0.984718	1.000000	1.000000	BFGS 28	SOS	Exponential	Exponential	33.61	175.52	5.67%	10.25%	237.16	0.0039
60	MLP 1-7-1	0.979533	1.000000	1.000000	BFGS 7	SOS	Logistics	Exponential	137.72	301.82	15.77%	20.11%	440.74	0.1426
60	MLP 1-4-1	0.985228	1.000000	1.000000	BFGS 18	SOS	Tanh	Tanh	24.83	134.35	1.70%	7.15%	186.08	0.1489
60	MLP 1-4-1	0.985445	1.000000	1.000000	BFGS 22	SOS	Exponential	Tanh	15.91	132.92	2.92%	6.56%	178.70	0.0044
80	MLP 1-5-1	0.984004	1.000000	1.000000	BFGS 25	SOS	Exponential	Logistics	42.53	185.12	7.07%	11.52%	266.63	0.0035
80	MLP 1-3-1	0.986307	1.000000	1.000000	BFGS 8	SOS	Linear	Linear	10.70	130.48	1.10%	5.76%	163.79	0.0039
80	MLP 1-7-1	0.986307	1.000000	1.000000	BFGS 6	SOS	Linear	Linear	10.11	130.12	1.06%	5.73%	163.63	0.1426

https://doi.org/10.55343/CogSust.102

Network number	Network name	Quality (training)	Quality (testing)	Quality (validation)	Learning algorithm	Error function	Activation (hidden)	Activation (output)	Errors					
									ME	MAE	MPE	MAPE	SSE	Theil
80	MLP 1-3-1	0.983467	1.000000	1.000000	BFGS 18	SOS	Logistics	Tanh	15.85	132.93	1.79%	6.60%	191.52	0.1489
80	MLP 1-2-1	0.985699	1.000000	1.000000	BFGS 88	SOS	Tanh	Tanh	18.95	116.26	1.01%	5.28%	175.44	0.0044
100	MLP 1-2-1	0.980835	1.000000	1.000000	BFGS 4	SOS	Exponential	Tanh	152.57	261.74	14.98%	17.76%	389.68	0.0035
100	MLP 1-5-1	0.981812	1.000000	1.000000	BFGS 14	SOS	Logistics	Logistics	77.48	230.84	10.33%	14.48%	328.62	0.0039
100	MLP 1-5-1	0.981972	1.000000	1.000000	BFGS 5	SOS	Exponential	Linear	32.76	193.92	6.86%	11.39%	273.65	0.1426
100	MLP 1-8-1	0.975335	1.000000	1.000000	BFGS 5	SOS	Linear	Tanh	13.48	214.31	5.45%	11.51%	283.43	0.1489
100	MLP 1-5-1	0.981904	1.000000	1.000000	BFGS 14	SOS	Logistics	Exponential	67.07	224.30	9.85%	14.64%	328.61	0.0044
200	MLP 1-7-1	0.982855	1.000000	1.000000	BFGS 11	SOS	Logistics	Exponential	78.88	211.08	9.66%	13.70%	303.74	0.0035
200	MLP 1-2-1	0.984554	1.000000	1.000000	BFGS 38	SOS	Exponential	Exponential	53.59	184.81	7.27%	11.34%	256.84	0.0039
200	MLP 1-3-1	0.984417	1.000000	1.000000	BFGS 76	SOS	Tanh	Logistics	27.98	88.21	0.75%	3.42%	147.43	0.1426
200	MLP 1-8-1	0.986698	1.000000	1.000000	BFGS 32	SOS	Tanh	Linear	20.01	108.99	0.42%	4.41%	159.41	0.1489
200	MLP 1-5-1	0.975155	1.000000	1.000000	BFGS 4	SOS	Linear	Tanh	19.52	204.69	3.79%	10.75%	283.68	0.0044
								Minimal	**0.61**	**88.21**	**0.07%**	**3.42%**	**140.34**	**0.0035**

https://doi.org/10.55343/CogSust.102

It is possible to conclude that based on the test findings shown, the number of pedestrian-related traffic incidents will stabilise in the upcoming years, almost regardless of the random sample size used. The average percentage error is reduced when the proportion of the training set is higher than that of the test and validation sets. The error was 3.70% for the training set (70%), 3.42% for the test set (15%), and 15% for the validation test in the proportions (70 : 15 : 15). The findings are impacted, on the one hand, by the rising number of automobiles on Polish roads and the recent epidemic (Fig. 3).

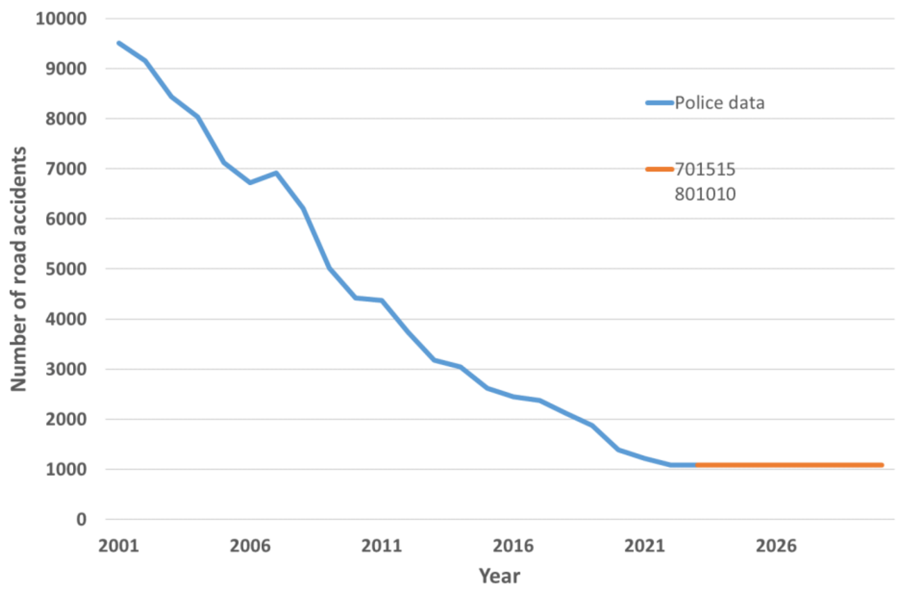

Figure 3. Projected number of road accidents for 2024-2030

4. Conclusion

The frequency of accidents caused by pedestrians was predicted using neural networks, and the study was conducted in the *Statistica* environment. The computer evaluated the study's weights in a way that minimised both the mean absolute error and the mean absolute percentage error.

Based on the data collected, we may still anticipate stabilising the number of traffic accidents. On the one hand, the recent epidemic and the growing number of cars on the road impact this. The computed forecast errors show how accurate the models were.

Measures to further reduce the number of traffic accidents should be implemented in light of the forecasts that have been obtained. One solution could be increased penalties for traffic offences on Polish roads, effective January 1, 2022. Undoubtedly, the pandemic affected the research findings that were acquired since it drastically altered the number of traffic accidents.

In their future study, the authors want to employ other statistical techniques and account for additional elements impacting the frequency of accidents to ascertain the total number of traffic accidents. These might include the amount of traffic, the kind of weather, the driver's age who caused the collision, or exponential techniques for calculating the frequency of traffic accidents.

References

Abdullah, E., Emam, A. (2017). Traffic accidents analyser using big data. *2015 International Conference on Computational Science and Computational Intelligence (CSCI).* 392–397. DOI: https://doi.org/ms4z

Al-Madani, H. (2018). Global road fatality trends' estimates based on country-wise microlevel data. *Accident Analysis & Preview*. 111, 297–310, DOI: https://doi.org/gcz5cx

Arteaga, C., Paz, A., Park, J. (2020). Injury severity on traffic crashes: A text mining with an interpretable machine-learning approach. *Safety Science*. 132 (2020), Article 104988. DOI: https://doi.org/gm8njj

Bąk. I., Cheba, K., Szczecińska, B. (2019). The statistical analysis of road traffic in cities of Poland. *Transportation Research Procedia*. 39. 14–23. DOI: https://doi.org/ms4v

https://doi.org/10.55343/CogSust.102

Baranyai, D., Sipos, T. (2022). Black-spot analysis in Hungary based on kernel density estimation. *Sustainability*. 14(14), 8335. DOI: https://doi.org/jswq

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)*. 1–5. DOI: https://doi.org/mt6b

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. DOI: https://doi.org/gkcnd5

Chen, C. (2017). Analysis and forecast of traffic accident big data. *ITM Web Conference*. 12, 04029. DOI: https://doi.org/ms4w

Chudy-Laskowska, K., Pisula, T. (2015). Prognozowanie liczby wypadków drogowych na Podkarpaciu [Forecasting the number of road accidents in Subcarpathia]. *Logistyka*. 2015(4), 2782–2796. URL: https://www.researchgate.net/publication/316596821_Prognozowanie_liczby_wypadkow_drogowych_na_Podkarpaciu

Data mining techniques – Techniki zgłębiania danych (data mining) (n. d.). StatSoft. URL: https://www.statsoft.pl/textbook/stathome_stat.html?https%3A%2F%2Fwww.statsoft.pl%2Ftextbook%2Fstdatmin.html (Downloaded: 2 May 2022 10:43)

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. URL: https://connectjournals.com/03899.2020.16.607

Forecasting based on time series – Prognozowanie na podstawie szeregów czasowych (n. d.). URL: http://pis.rezolwenta.eu.org/Materialy/PiS-W-5.pdf (Downloaded: 2 May 2022 10:43)

Gorzelanczyk, P., Pyszewska, D., Kalina, T., Jurkovic, M. (2020). Analysis of road traffic safety in the Piła Poviat. *Scientific Journal of Silesian University of Technology. Series Transport*. 107, 33–52. DOI: https://doi.org/ms4x

Helgason, A. (2016). Fractional integration methods and short time series: evidence from a simulation study. *Political Analysis*. 24(1), 59–68. URL: http://www.jstor.org/stable/24573204

Karlaftis M., Vlahogianni E. (2009). Memory properties and fractional integration in transportation time-series. *Transportation Research, Part C: Emerging Technologies*. 17(4), 444–453. DOI: https://doi.org/c3x6wb

Kashpruk, N. (2020). Comparative research of statistical models and soft computing for identification of time series and forecasting. Opole University of Technology. URL: https://dbc.wroc.pl/Content/108023/Praca%20Doktorska_%20Nataliia%20Kashpruk_popr.pdf

Khaliq, K. A., Chughtai, O., Shahwani, A., Qayyum, A., Pannek, J. (2019). Road accidents detection, data collection and data analysis using V2X communication and edge/cloud computing. *Electronics*. 8(8). DOI: https://doi.org/ms42

Kumar, S., Viswanadham, V., Bharathi, B. (2019). Analysis of road accident. *IOP Conference Series Materials Science and Engineering*. 590(1): 012029. DOI: https://doi.org/ms43

Lavrenz, S., Vlahogianni, E., Gkritza, K., Ke, Y. (2018). Time series modeling in traffic safety research. *Accident Analysis & Prevention*. 117, 368–380. DOI: https://doi.org/gdvwn3

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, e253. DOI: https://doi.org/mzpk

Łobejko, S. (ed.) (2015). Time series analysis and forecasting with SAS. Publisher: SGH, Warszawa.

Mamczur, M. (2020). *Jak działa regresja liniowa? I czy warto ją stosować?* [How does linear regression work? And is it worth using?]. URL: https://miroslawmamczur.pl/jak-dziala-regresja-liniowa-i-czy-warto-ja-stosowac/ (Downloaded: 2 May 2022 10:43)

McIlroy, R. C., Plant, K. A., Hoque, M. S., Wu J., Kokwaro, G. O., Nam, V. H., Stanton, N. A. (2019). Who is responsible for global road safety? A cross-cultural comparison of Actor Maps. *Accident Analysis & Prevention*. 122, 8–18. DOI: https://doi.org/gfpgvt

Mekonnen, A. A., Beza, A. D., Sipos, T. (2022). Estimating the Value of Statistical Life in a Road Safety Context Based on the Contingent Valuation Method. *Journal of Advanced Transportation*, Article 3047794. DOI: https://doi.org/kc3r

Monedero B. D., Gil-Alana L. A., Martíneza M. C. V. (2021).Road accidents in Spain: Are they persistent? *IATSS Research*. 45(3), 317–325. DOI: https://doi.org/ms44

Oronowicz-Jaśkowiak W. (2019). The application of neural networks in the work of forensic experts in child abuse cases. *Advances in Psychiatry and Neurology*. 28(4), 273–282. DOI: https://doi.org/ms45

Perczak G., Fiszeder P. (2014). Model GARCH− wykorzystanie dodatkowych informacji o cenach minimalnych i maksymalnych [GARCH model – using additional information on minimum and maximum prices]. *Bank i Kredyt*. 45(2), 105–132.

Piłatowska M. (2012). Wybór rzędu autoregresji w zależności od parametrów modelu generującego [The choice of the order of autoregression depending on the parameters of the generating model]. *Ekonometria*. 4(38), 16–35. URL: https://dbc.wroc.pl/Content/22753/Pilatowska_Wybor_Rzedu_Autoregresji_w_Zale%C5%BCnosci_Od_Parametrow.pdf

https://doi.org/10.55343/CogSust.102

Poland Number of Registered Vehicles URL: https://www.ceicdata.com/en/indicator/poland/number-of-registered-vehicles (Downloaded: 28 May 2024 15:43)

Poland Population URL: https://www.worldometers.info/world-population/poland-population/ (Downloaded: 28 May 2024 15:43)

Prochazka J., Camaj M. (2017). Modelling the number of road accidents of uninsured drivers and their severity. *Proceedings of International Academic Conferences*. 5408040. URL: https://ideas.repec.org/p/sek/iacpro/5408040.html

Procházka J., Flimmel S., Čamaj M., Bašta M. (2017). Modelling the Number of Road Accidents. *20-th AMSE. Applications of Mathematics and Statistics in Economics. International Scientific Conference*. *Proceedings – Full Text Papers*. 355–364. DOI: https://doi.org/ms46

Rajput H., Som, T., Kar, S. (2015). An automated vehicle license plate recognition system. *Computer*. 48(8), 56–61. DOI: https://doi.org/ms47

Random forest – Las losowy (n. d.) URL: https://pl.wikipedia.org/wiki/Las_losowy (Downloaded: 2 May 2022 10:43)

Road Accident Statistics (2024). *Statystyka*. URL: https://statystyka.policja.pl (Downloaded: 2 May 2024 10:43)

Road safety statistics in the EU, (2024) URL: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Road_safety_statistics_in_the_EU&oldid=630784 (Downloaded: 28 May 2024 15:43)

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. 2018 International Conference on Data Science and Engineering (ICDSE). DOI: https://doi.org/ms48

Szmuksta-Zawadzka M., Zawadzki J. (2009). Forecasting on the basis of Holt-Winters models for complete and incomplete data. *Research papers of the Wrocław University of Economics*. 38. URL: https://www.dbc.wroc.pl/Content/15648/Szmuksta-Zawadzka_Zawadzka_O_prognozowaniu_na_podstawie_modeli.pdf

Tambouratzis T., Souliou D., Chalikias M., Gregoriades A. (2014). 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*. 4(1), 31–42. DOI: https://doi.org/ms49

Top Advantages and Disadvantages of Hadoop 3 (n. d.) DataFlair URL: https://data-flair.training/blogs/advantages-and-disadvantages-of-hadoop/ (Downloaded: 2 May 2022 10:43)

Vilaça M., Silva N., Coelho M. C. (2017). Statistical analysis of the occurrence and severity of crashes involving vulnerable road users. *Transport Research Procedia*. 27, 1113–1120, DOI: https://doi.org/ms5b

WHO (2018). *The global status on road safety*. URL: https://www.who.int/publications/i/item/9789241565684 (Downloaded: 2 May 2022 10:43)

Wójcik A. (2014). *Autoregressive vector models as a response to the critique of multi-equation structural econometric models.* Publishing House of the University of Economics in Katowice. Vol. 193.

Wrobel M. S. (2011). *Application of neural fuzzy systems in chemistry*. PhD thesis. University of Silesia, Katowice. URL: https://rebus.us.edu.pl/bitstream/20.500.12128/5266/1/Wrobel_Zastosowanie_neuronowych_systemow.pdf

Yang Z. Zhang W., Feng J. Predicting multiple types of traffic accident severity with explanations: A multi-task deep learning framework. *Safety Science*. 146, 105522. DOI: https://doi.org/gnxr67

Zheng Z., Wang C., Wang P., Xiong Y., Zhang F., Lv Y. (2018). Framework for fusing traffic information from social and physical transportation data. *PLoS One*, 13. DOI: https://doi.org/gd2dm2

Zhu L., Lu L., Zhang W., Zhao Y., Song M. (2019). Analysis of accident severity for curved roadways based on Bayesian networks. *Sustainability*. 11(8), 2223. DOI: https://doi.org/ms5c