

The Eurasia Proceedings of Science, Technology, Engineering and Mathematics (EPSTEM), 2026

Volume 39, Pages 57-71

IConTech 2026: International Conference on Technology

Prediction of Accidents Using Artificial Neural Networks in North Cyprus

Hillary Makonise

Near East University

Mustafa Alas

Near East University

Abstract: When compared to global traffic catastrophes, North Cyprus road system ranks among the safest up to the current fiscal quarter of 2025. Considering an approximate estimate of about 1.3 million deaths and millions of non-fatal injuries which are caused by road accidents each year, making them a serious public health concern, it is challenging to anticipate and avoid traffic accidents on new road segments due to a multitude of subjective and objective factors. With minimal data sets at disposal, artificial neural networks (ANN) have proven to be successful in forecasting traffic accidents. Therefore, it is crucial to properly evaluate the pertinent hazards. A methodology for creating an accident risk prediction model that can be utilized as a tool for infrastructure management decision-making is presented in this study. These models use easily determined objective factors, such as road length, terrain type, road width, average daily traffic volume, and speed limit. The approach makes it possible to handle the supplied data appropriately, looks at how it can be utilized to create models using artificial neural networks (ANNs), and establishes a methodical ANN optimization procedure to find the best ANN model architecture. The approach is applied using accident count data as a result, it is essential to accurately assess the relevant risks. This paper presents an approach for developing an accident risk prediction model that can be used as a tool for infrastructure management decision-making. The method offers a systematic ANN optimization process to determine the optimal ANN model architecture, making it feasible to handle the provided data effectively, and examines how it may be used to develop models using artificial neural networks (ANNs). The method is used with data on the number of accidents on North Cyprus road.

Keywords: Artificial neural network, Traffic accident, Infrastructure planning

Introduction

One of the main causes of death globally is traffic accidents, approximately 50 million people sustain non-fatal injuries and nearly 1.19 million people die in traffic accidents, according to the World Health Organization (WHO, 2025). Road traffic crashes are the primary cause of death for populations of moderate age (WHO, 2025). The significant issue is that 93% of deaths occur in developing nations with low and moderate incomes (WHO, 2025). In addition to causing human suffering, traffic accidents have a substantial economic impact, costing nations up to 3% of their GDP (WHO). Numerous factors can be linked to the frequency of traffic accidents. While some of these characteristics are quantitative, such as road volume (annual average daily traffic Algorithms 2024), others are subjective, like driver knowledge, training level, experience, the impact of drinking, drugs. Road geometry (curvature, slopes, lane width, shoulder width), road type (freeway or two-lane road), road conditions (pavement quality and potential pavement surface damage), weather (wind, ice, snow, rain), vehicle maintenance, speed limits, frequency of police checks. Increased traffic volume on nearly all highways' results from an increase in the number of vehicles on the road, which raises the likelihood of vehicle collisions. However, the type of road, road geometry, and AADT (which will probably rise in the future) are acknowledged as objective characteristics that are difficult to modify in the future (Federal Highway Administration). Despite the fact that accidents are random and unpredictable at micro-level, statistical models can predict reliable estimates of expected accidents by relating aggregates of accidents to various explanatory

- This is an Open Access article distributed under the terms of the Creative Commons Attribution-Noncommercial 4.0 Unported License, permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

- Selection and peer-review under responsibility of the Organizing Committee of the Conference

© 2026 Published by ISRES Publishing: www.isres.org

measures of flow, site characteristics, and road geometry at macro-level. There are other variables that can be altered, such as the ability to repair pavement, the daily fluctuations in the weather, the improvement of car upkeep through regular inspections, the ability to modify speed restrictions, etc. Many scientific fields make substantial use of traditional machine learning models, including artificial neural networks (ANN), random forest regression (RFR), support vector machines (SVM), extreme learning machines (ELM), K-nearest neighbors (KNN), and decision trees (DT). Based on statistical learning theory, the SVM is a popular discriminant method that is well regarded for its capacity for generalization. By investigating the balance between the model's complexity and the training error, the ideal network is found (Vapnik, 1995; Cortes & Vapnik, 1995). By creating the input weights and biases of the hidden layers at random, the ELM creates a single-layer feedforward network (Huang et al., 2006). A wide range of cutting-edge machine learning methods, including ensemble learning models like XGBoost, LightGBM, and CatBoost, are appropriate for sequence data. The benefits of the XGBoost model are particularly evident in its excellent prediction accuracy and interpretability (Chen & Guestrin, 2016; Ke et al., 2017; Prokhorenkova et al., 2018).

Large data sets and GPU training are made possible by the LightGBM model. It has been demonstrated that the LightGBM models outperform XGBoost in terms of accuracy and speed. Additionally, the integration of gradient boosting-based category features enabled by the CatBoost algorithm indicates that data fusion allows for greater forecasting accuracy (Ke et al., 2017, Ke et al., 2017; Chen & Guestrin, 2016). In recent years, the use of artificial neural networks (ANN) has demonstrated its viability by forecasting and displaying desirable outcomes despite small data sets (Jain et al., 2015, Zhang et al., 2025, Li et al., 2024). For the enhancement of traffic safety management, the accuracy and outcomes of traffic accident prediction models are crucial.

The results show that the most significant factors which increase traffic accidents on urban roads are highway width, head-on collisions, the type of vehicle at fault, ignoring lateral clearance, following distance, inability to control the vehicle, over speeding, and driver deviation to the left. This study offers artificial neural network (ANN) model for evaluating traffic accidents on state roads in KKTC.

Objectives

- i.** To develop an artificial neural network (ANN) based predictive model capable of accurately forecasting accident occurrence using historical accident, traffic, environmental, and roadway data.
- ii.** To identify and quantify the most influential factors contributing to accidents by analyzing input features (e.g., traffic volume, weather conditions, road geometry, human factors) within the ANN framework.
- iii.** To evaluate and compare the performance of different ANN architectures and training algorithms in terms of prediction accuracy, robustness, and generalization capability.
- iv.** To validate the proposed ANN model using real-world datasets and benchmark its performance against traditional statistical and machine-learning accident prediction methods.
- v.** To demonstrate the practical applicability of the proposed prediction model for proactive accident prevention, decision support, and traffic safety planning by transportation authorities

Research Questions

1. How accurately can an artificial neural network predict traffic accident occurrence in North Cyprus based on historical accident data?
2. Which factors (e.g., time of day, weather conditions, road type, traffic volume, driver demographics) contribute most significantly to accident prediction when using an ANN?
3. How does the performance of an artificial neural network compare with traditional statistical models (e.g., logistic regression) for predicting traffic accidents in North Cyprus?
4. Can an ANN effectively identify high-risk locations and time periods for traffic accidents in North Cyprus?
5. How does varying the architecture of the artificial neural network affect accident prediction accuracy?

Hypothesis

H₁ When trained on the same historical accident and traffic datasets, artificial neural network-based models predict accident occurrence and severity with far higher accuracy than conventional statistical and machine-learning models.

H₂ When compared to linear modeling techniques, artificial neural networks greatly enhance accident prediction performance by capturing intricate nonlinear interactions between traffic, environmental, and highway factors.

Literature Review

Numerous studies globally have focused on predicting traffic accidents and comprehending the factors which compromises traffic safety. These studies have investigated the inter-connectivity of variables using a variety of methods and methodologies (Lord & Mannering, 2010; Mannering & Bhat, 2014). Chapter 10 of the Highway Safety Manual specifically discusses the prediction model for rural two-lane, two-way highways, highlighting the influence of traffic volume (AADT) via project geometry and safety performance functions (SPF) and crash modification factors (American Association of State Highway and Transportation Officials (AASHTO, 2010).

In a bid to examine the geometric features leading to accidents, Bared and Vogt (1999) investigated Minnesota and Washington, two lane roads both urban and rural set ups, aiming to find the parameters that has the biggest impact on the frequency of traffic accidents. Despite being restricted to two-lane roads, Persaud presented one of the first studies for performing independent evaluations for curves and tangents. The reliant Road geometry and traffic flow were the independent variables, and crash frequency was the dependent variable. Generalized linear modeling was used to calibrate the regression models. Additionally, a dummy variable for "flat" or "undulating" topography was employed. AADT, section length (L), and curvature (1/R) were found to enhance crash frequency for curves. For tangents, the annual number of accidents rises with AADT and L. It has been demonstrated that accident frequencies are higher on undulating terrain than on flat road segments. Using crash data for rural two-lane highways in Minnesota et al. (2010) found that annual average daily traffic (AADT), percentage of trucks, and shoulder width significantly influenced the occurrence of head-on collisions (Geedipally et al., 2010). In Portugal, Cardoso (2018) developed an accident prediction model for curves and tangents on two-lane roadways (Cardoso, 2018). Abdel-Aty and Radwan (2000) predicted crash frequency as a function of AADT, degree of horizontal curvature, section length, lane width, median width, shoulder width, and urban-rural classification using a negative binomial distribution (Abdel-Aty & Radwan, 2000). The present research established accident prediction models for both tangents and curves, as well as for road sections with paved and unpaved shoulders (Cardoso, 2018). Harwood and colleagues (2000) focused on two-lane rural highways and developed a safety performance prediction algorithm (Harwood et al., 2000). Mayora et al. (2016) analyzed approximately 3,450 km of two-lane rural roads over a five-year period in the Valencia and West Castile regions of Spain (Mayora et al., 2016).

About 170 km of two-lane rural roads in Italy were examined over a five-year period in the research by Cafiso et al. (2010). Variables included curvature (radius, length), tangent length, and cross-section elements (lane width and shoulder width and type). Ackaah and Salifu's paper (2021) provided information on traffic accidents in Ghana. The Generalized Linear Model (GLM) with a Negative Binomial error structure was the prediction model created in this study. Dinu and Veeraragavan (2011) created a random parameter model for accident prediction in India. The model's variables were shoulder width and curvature (both horizontal and vertical), length (km), the percentage of buses, trucks, autos, two-wheelers, and traffic volume. Turner and et al. (2016) created models for predicting accidents on two-lane rural roads in New Zealand. Analysis was done on 6200 km of state roadways. Numerous factors, including traffic volume, road geometry, cross-section, road surfacing, roadside hazards, and driveway density, were linked to a number of accidents by the proposed model.

Numerous linear models were created for various accident situations. Three distinct statistical techniques were illustrated in Deublein et al. (2013). The models employed were: (1) gamma-updating of the rates of road user injuries and injury accidents, (2) hierarchical multivariate (3) Bayesian inference techniques and Poisson-lognormal regression analysis. Impact Traffic volume, the proportion of trucks and buses, speed, curvature, and the quantity of lanes. To help traffic systems, several machine learning techniques are tried. LSTM/RNN is frequently utilized for sequence modeling in transportation planning. Prediction (Deublein et al., 2013), CNN are used for processing spatial and visual information, such as image or video-based traffic scene analysis, and hybrid models combining CNNs with LSTMs have been applied to enhance prediction performance by leveraging both temporal patterns and learned features (Liu, 2022; Zheng, 2019).

Hosseinpour et al. (2016) analyzed 200 Malaysian state road segments over five years, examining horizontal curvature, terrain type, heavy vehicle percentage, and access locations, and noted that conventional crash prediction methods, such as generalized linear regression, fail to account for multilevel data structures. In Turkey, Çodur and Tortum (2015) demonstrated the application of Artificial Neural Networks (ANN) for accident prediction, while Chang (2005) applied ANN models to Taiwan's national freeway system. Xiaoxia, Chen and Liang (2018) combined Bayesian Networks with Rough Sets to analyze traffic accidents, and Olmuş

and Erbaş (2012) employed Log-Linear Models for incident evaluation. Marković et al. (2019) utilized Bayesian Neural Networks to identify factors influencing traffic accidents, whereas Milenković et al. (2019) examined the impact of road and traffic characteristics on fatal accident frequency in Serbia using regression analysis, marking the first study of its kind in the country. Tubić et al. Graovac (2019) calculated the costs per kilometer associated with traffic accidents on major Serbian roads. In Turkey, Akgüngör and Doğan (2009) developed accident prediction models using ANN, non-linear regression, and hybrid ANN–genetic algorithm approaches, including evaluations based on the Modified Smeed and Adapted Andreassen methods. Similarly, Cansız (2011) applied ANN and non-linear models using the Smeed equation to estimate fatalities in traffic accidents

Broadly, a lot of study has been done on the effects of various elements on traffic safety and the frequency of traffic accidents. Many of these studies have concentrated on the impact of traffic volume, project geometry, traffic management features, and other related aspects. Numerous accident prediction models and techniques for evaluating and forecasting traffic accidents have been developed as a result of these investigations. Our knowledge of the variables influencing traffic accidents has increased as a result of these studies, each of which used particular instruments and techniques. Nonetheless, it is necessary to look at the use of objective elements for accident prediction in the Republic of Northern Cyprus. This study attempts to effectively anticipate traffic accidents and assess their severity based on criteria using Artificial Neural Network models. The findings from this research will provide insights for improving road safety and supporting decision-making in transportation planning.

Methodology

A comprehensive quantitative investigation of the factors influencing traffic accident frequency was conducted using roadway and accident data obtained from official sources in the Republic of Northern Cyprus. Annual Average Daily Traffic (AADT) data were collected from automatic traffic counters, while accident records for the national highway network were obtained from the Directorate of Traffic for the period 2020–2025. In total, 1,580 detailed accident reports were examined. Each report contained information on the date and location of the accident, pavement and vehicle types, driver characteristics, roadway and environmental conditions, time of occurrence, lighting conditions, and accident severity indicators, including fatalities, injuries, and the number of vehicles involved. Additional data describing roadway geometric and operational characteristics, such as horizontal and vertical curvature, lane configuration, median width, shoulder width, and traffic volume, were obtained from Traffic Directorate of Police bulletins. Based on relevance to accident occurrence and data completeness, the collected information was organized into eighteen variables, as presented in Table 1. To ensure that the analysis captured a wide range of traffic and geometric conditions, major highways exhibiting varying operational characteristics were selected and divided into segments with relatively homogeneous features as shown on Table 1.

Table 1. Summary of input

Design Parameters	Classification	Binary Numerical code
X ₁	Years	2020-2025
X ₂	Highway Sections	Km
X ₃	Drivers' gender	Male/ Female
X ₄	Time (day/night)	Daylight, night with/out illumination
X ₅	Road Surface Condition- RSC	Dry, wet, snow, ice
X ₆	Shoulder Width (SW)	1.2m-3m
X ₇	Median Width (MW)	3m-6m,6m-12m
X ₈	Highway Gradient (HG)	3%-6%
X ₉	Horizontal Curvature (HC)	IRC-38
X ₁₀	Vertical Curvature (VC)	IRC-38
X ₁₁	Annual Average Daily Traffic (AADT)	Numerical value
X ₁₂	Heavy Vehicle Traffic Accidents	Numerical value
X ₁₃	Light Vehicle Traffic Accidents	Numerical value
X ₁₄	Summer Accidents	Numerical value
X ₁₅	Winter Accidents	Numerical value
X ₁₆	Fall Accidents	Numerical value
X ₁₇	Spring Accidents	Numerical value
Y	Accumulative Accidents Value	Numerical value

Table 2. Roads classification

Mark	Highway name	Type	Length
Road 1	Lefkosia-Magusa	Four lane state road Class 1	61 km
Road 2	Lefkosia-Girne	Four lane state road Class 2	23km
Road 3	Lefkosia-Guzelyurt	Four lane state road Class 3	45km
Road 4	Lefkosia – Esentepe	Two lane state road Class 4	57km

The highways were identified as suitable study corridors due to their functional importance and variation in traffic demand and roadway design. Segmenting the highways in this manner reduced internal variability and improved the statistical consistency of the modeling process. Figure 1 presents Data Landscape from 2020 to 2025, highlighting four major road corridors extending from Lefkosia to Magusa, Girne, Guzelyurt, and Esentepe. Each route is classified by distance and road type, allowing us to compare different traffic conditions. Over this five-year period, we gathered 1,580 detailed accident reports, providing a comprehensive foundation for analyzing patterns, identifying high-risk zones, and informing data-driven strategies to enhance road safety across Cyprus.

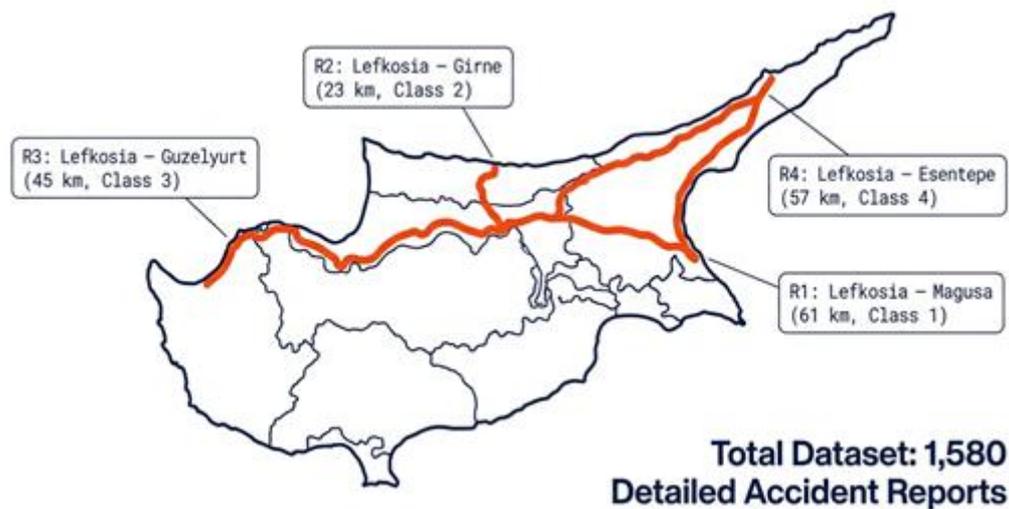


Figure 1. Highway sections in North Cyprus

Figure 1 illustrates the road networks for the republic of Northern Cyprus. The accumulative accident figure was extracted from published Traffic Agency data in Kibris Gazetesi, together with the geo location of the exact place where it occurred. Within the Ayuda Neural Intelligence framework, the dataset was structured to support artificial neural network (ANN) modeling and comparative machine learning analysis. Variables representing traffic exposure, roadway characteristics, and accident composition, including year, road length, lane configuration, AADT, heavy accident counts, and light accident counts, were treated as model inputs. These variables are widely recognized in traffic safety research as significant contributors to accident occurrence and provide a balanced representation of temporal, structural, and operational influences. Road identification codes were retained for descriptive and segmentation purposes only and were excluded from numerical model training to prevent categorical bias. The dependent variable was defined as accidents per kilometer, calculated by normalizing total accident counts by segment length, thereby allowing meaningful comparison across roadway segments of different sizes and reducing exposure-related bias.

Prior to model development, all numerical input variables were standardized using z-score normalization to ensure comparability across differing measurement scales and to enhance numerical stability during training. The normalized dataset was subsequently divided into training and testing subsets to evaluate model generalization performance. An artificial neural network model was then developed within the Ayuda Neural Intelligence framework, with the network architecture optimized through systematic adjustment of hidden layers and neuron counts to minimize prediction error. Model training employed iterative learning to capture nonlinear relationships between traffic exposure, roadway characteristics, and accident frequency. For benchmarking purposes, multiple linear regression, support vector regression, and random forest models were also developed

using the same input–output structure and data partitions. Model performance was evaluated using mean squared error, root mean squared error, and the coefficient of determination (R^2), allowing a consistent and objective comparison between the ANN and conventional machine learning approaches. Ayuda Neural Intelligence graph is shown below on figure 2;

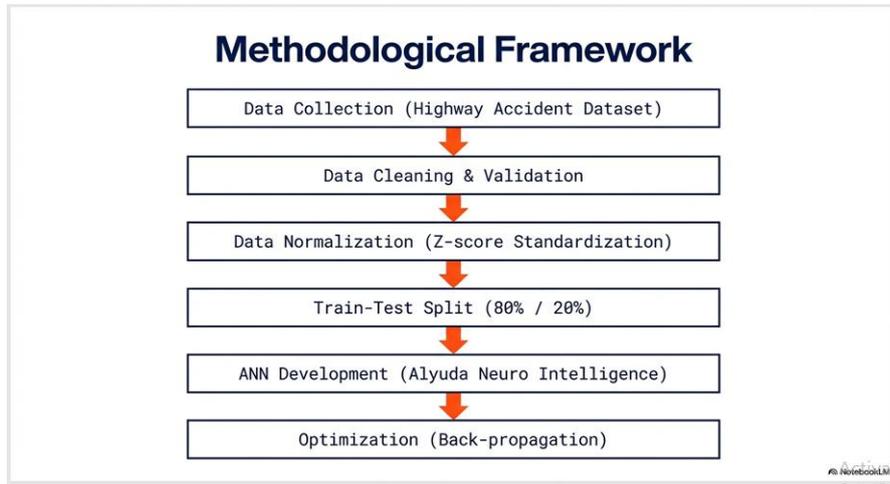


Figure 2. Framework for the ANN architecture

Non-linear system modeling is considerably more challenging. Three layers which constitute a multilayer perceptron (MLP) structural model:(input, hidden, and output) were used to create artificial neural network models (ANN) for predicting the number of traffic accidents based on road conditions. Neural networks can be used to identify patterns and trends that are too complex for humans or other computer methods to recognize because of their stupendous ability to learn complex relationships from imprecise input. Artificial neural networks can be used in many powerful ways, such as learning and reproducing rules or operations from examples, analyzing and generalizing from samples, making predictions or memorizing features of given data, and matching or forming associations between new and old data .Network training is the process by which the environment in which the network is placed continuously simulates the connection weights and biases of ANN. Finding a set of connection strengths and biases that lead an ANN to generate outputs that are equal to or near targets is the main objective of training in order to minimize an error function. To put it another way, training seeks to estimate the parameters (W_1 , W_2 , b_1 , and b_2) by minimizing an error function, like the output values' mean square error (MSE), which is stated as equation 1:

$$MSE = \sum_{i=k}^N \left(\frac{t-a}{N} \right) \quad N = \text{number of data} \quad (1)$$

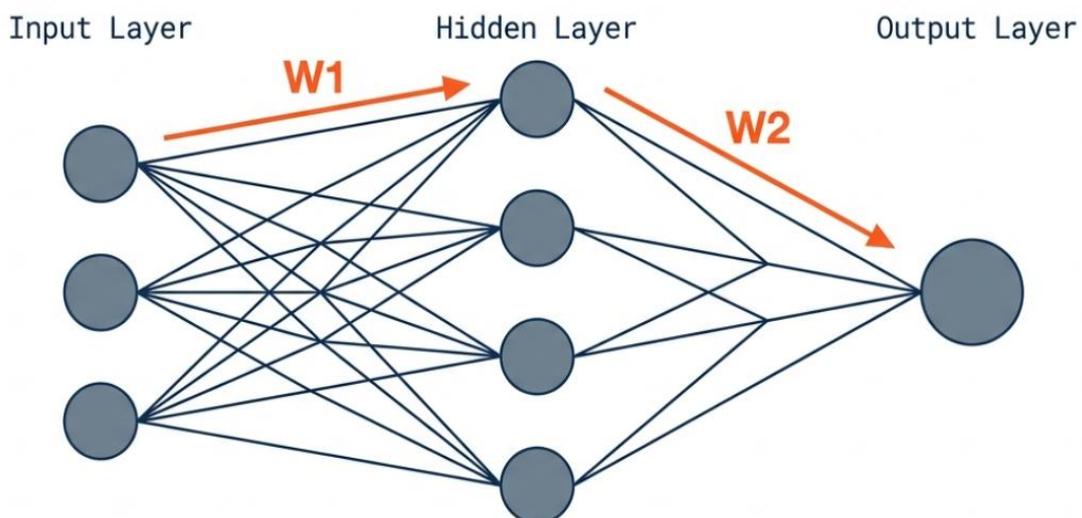


Figure 3. Structure of ANN

For several decades, researchers have investigated nonlinear system modeling and have demonstrated that artificial neural networks (ANNs) are capable of learning nonlinear system dynamics without requiring explicit mathematical models (Haykin, 1999; Rumelhart, Hinton, & Williams, 1986). Artificial neural networks have been successfully applied to a wide range of tasks, including learning and reproducing rules or processes from example data, evaluating and extrapolating from samples, forecasting or retaining characteristics of input data, and identifying relationships between new and previously learned data (Bishop, 1995; Haykin, 1999).

As illustrated in Figure 3, the network architecture consists of multiple nodes in the input layer and a single node in the output layer, commonly referred to as the target value. Strong approximation capability is achieved by employing a linear activation function in the output layer and a differentiable sigmoid activation function in the hidden (intermediate) layer. There is no fixed rule or theoretical framework for determining the optimal number of neurons in the hidden layer; instead, this number is typically selected through an iterative trial-and-error process based on network performance (Bishop, 1995; Haykin, 1999).

Development and Application of ANN Models

Several software programs are available for implementing the Levenberg–Marquardt algorithm; however, Alyuda Neuro Intelligence was selected as the most suitable platform for the artificial neural network (ANN) model investigation. Alyuda Neuro Intelligence is a professional neural network software designed to support the application of ANN techniques for solving real world problems, including forecasting, classification, and function approximation. The software incorporates intelligent features for dataset preprocessing, automated identification of efficient network architectures, performance analysis, and deployment of trained networks to new data. These capabilities enable experts to develop, test, and optimize neural network solutions more efficiently, thereby improving productivity and model performance (Alyuda Research, 2019).

In an ANN model, independent variables are referred to as inputs, while dependent variables are referred to as outputs. The input importance chart provided by Alyuda Neuro Intelligence illustrates the relative contribution of each input variable to the overall network performance. This chart allows users to identify and potentially eliminate less significant input variables without adversely affecting the predictive accuracy of the model. Understanding which inputs exert the greatest influence on network output is essential for model interpretation and simplification.

Input variable importance is calculated based on the decrease in network performance observed when a particular input is removed and excluded from the training process (Alyuda Research, 2019; Olden et al., 2004). Among the eighteen parameters initially considered for modeling, eight parameters were identified as significant according to the selected importance criteria. The relevant factors include:

- Years,
- Highway sections,
- Section length (km),
- Annual average daily traffic (AADT),
- The degree of horizontal curvature,
- The degree of vertical curvature,
- Traffic accidents with heavy vehicles (%),

Prior to network training, both input and output variables are normalized within the range 0-1 using a minimax technique. Columns for categories are automatically encoded using the Alyuda Neuro Intelligence Program's One-of-N technique during data preprocessing. A column with N different values is encoded into a collection of N numerical columns using the One-of-N encoding, with one column for every category in a metrics format. For instance, "Low" will be represented as {1,0,0}, "Medium" as {0,1,0}, and "High" as {0,0,1} in the Capacity column with the values "Low," "Medium," and "High." The minimum and maximum of the dataset were obtained and scaling factors selected so that these were transferred to the intended minimum and maximum values. The minimax algorithm is as follows:

$$H' = H \left(\frac{1}{H_2 - H_1} \right) - \left\langle \frac{H_1}{H_2 - H_1} \right\rangle H \quad (2)$$

Where:

H₂=Largest value of H
 H₁=smallest value of H
 H¹=unifying value of H

To have a significant increase in speed while minimizing network errors, Data normalization is very crucial. The efficacy of the back propagation training process depends on the number of neurons in the hidden layer; varied numbers of neurons within 1≤10 in the hidden layer were examined. Therefore, this study defines single output layer as the number of accidents for all roads. Therefore illustrated below are the highway section considered in this study table 4, and table 5 outlines the collected data set used for ML and ANN evaluation;

Table 3. Highway sections considered in the study public works department. (n.d.). *Road Network dataset* [Data set]. Cyprus National Access Point. https://www.traffic4cyprus.org.cy/en/dataset/road_network

Road ID	Highway Name	Lane Type	Road Class	Length (km)	Lane Code
R1	Lefkosia – Magusa	Four-lane state road	Class 1	61	2
R2	Lefkosia – Girne	Four-lane state road	Class 2	23	2
R3	Lefkosia – Guzelyurt	Four-lane state road	Class 3	45	2
R4	Lefkosia – Esentepe	Two-lane state road	Class 4	57	1

Overall, the dataset highlights variations in road classification and lane capacity, reflecting each highway’s relative importance, design standard, and role within the national transportation network with respect to the below used codes;

Lane Type Coding:

- Two-lane = 1
- Four-lane = 2

Table 4. 2020-2025 Accident data accumulated for ML

Year	Road ID	Length (km)	Lane Code	AADT	Heavy Acc.	Light Acc.	Total (Y)	Acc.	Accidents / km
2020	R1	61	2	36,200	38	102	140		2.30
2021	R1	61	2	37,500	40	108	148		2.43
2022	R1	61	2	38,900	41	112	153		2.51
2023	R1	61	2	40,200	42	118	160		2.62
2024	R1	61	2	41,300	44	121	165		2.70
2025	R1	61	2	42,800	46	125	171		2.80
2020	R2	23	2	29,800	22	64	86		3.74
2021	R2	23	2	30,500	24	68	92		4.00
2022	R2	23	2	31,600	25	71	96		4.17
2023	R2	23	2	32,400	27	74	101		4.39
2024	R2	23	2	33,200	28	76	104		4.52
2025	R2	23	2	34,000	30	79	109		4.74
2020	R3	45	2	25,900	29	83	112		2.49
2021	R3	45	2	26,800	31	86	117		2.60
2022	R3	45	2	27,900	33	89	122		2.71
2023	R3	45	2	28,700	35	92	127		2.82
2024	R3	45	2	29,500	36	95	131		2.91
2025	R3	45	2	30,300	38	98	136		3.02
2020	R4	57	1	18,400	26	91	117		2.05
2021	R4	57	1	19,200	28	95	123		2.16
2022	R4	57	1	20,100	30	99	129		2.26
2023	R4	57	1	21,000	32	103	135		2.37
2024	R4	57	1	21,800	34	107	141		2.47
2025	R4	57	1	22,600	36	111	147		2.58

Table 4 presents annual traffic and accident statistics for four major highways connecting Lefkosia to Magusa, Girne, Guzelyurt, and Esentepe from 2020 to 2025. Each record includes the road’s length, lane code, average annual daily traffic (AADT), and the number of accidents involving heavy and light vehicles, along with the total accidents and accident rate per kilometer. Overall, the results show a steady increase in traffic volumes across all routes over the six-year period. This rise in vehicle movement is accompanied by a gradual growth in total accidents and accident rates, suggesting a direct relationship between higher traffic demand and accident

occurrence. Among the four roads, the Lefkosia–Girne route (R2) stands out with the highest accident rate, despite being the shortest at 23 km. This indicates greater congestion or risk exposure compared to the other routes. The Lefkosia–Magusa (R1), Lefkosia–Guzelyurt (R3), and (Lefkosia–Esentepe R4) roads also show consistent year-on-year increases in both traffic flow and accidents, though at slightly lower rates.

In summary, the data reveal a pronounced and sustained escalation in both traffic density and accident incidence across Lefkosia’s principal arterial routes. This trend underscores the imperative for a comprehensive, evidence-based approach to transport policy encompassing advanced traffic management strategies, infrastructural modernization, and data-driven road safety interventions to mitigate the escalating mobility and safety challenges confronting the region’s transport network. Having the the accumulative collected data it would be mor logical to initiate the Machine Learning (ML) as demonstrated on the ANN learning Curve below on figure 4;

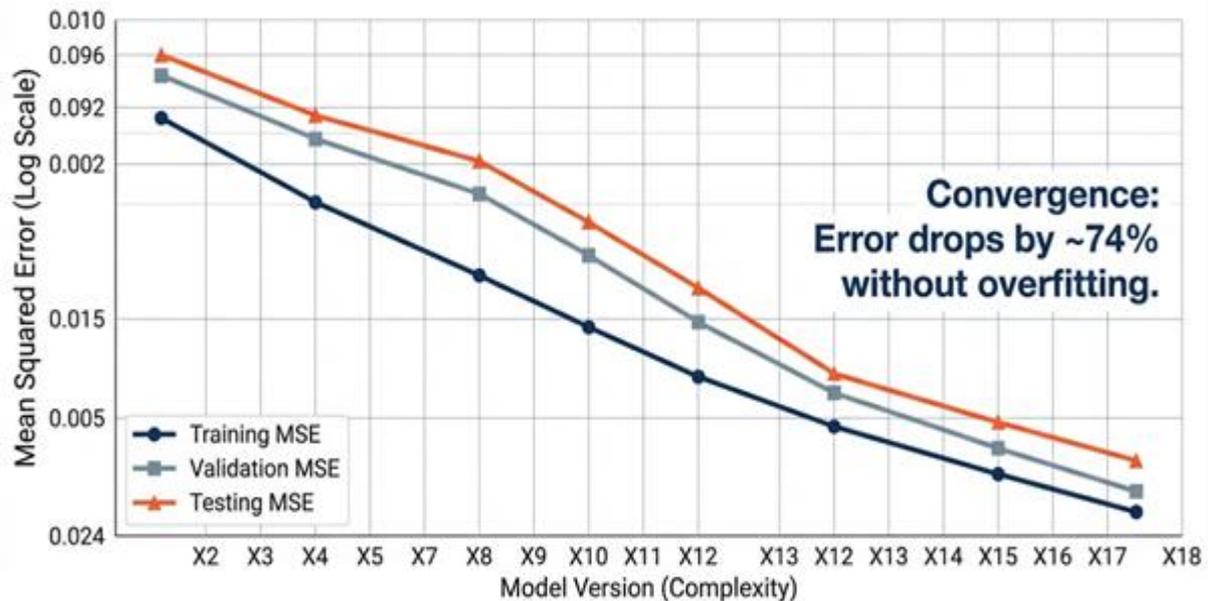


Figure 4. ANN learning curve

Figure 4 illustrates the training, validation, and testing mean squared error (MSE) of the Alyuda Neuro Intelligence model across 18 input scenarios, shown on a logarithmic scale. A clear and monotonic reduction in MSE is observed for all datasets, indicating progressive improvement in predictive performance and effective convergence of the learning process.

Quantitatively, the training MSE decreases from approximately 0.092 in Scenario 1 to 0.024 in Scenario 18, corresponding to an overall reduction of approximately 74%. Similarly, the validation MSE decreases from approximately 0.098 to 0.034, representing an improvement of about 65%, while the testing MSE decreases from approximately 0.101 to 0.039, yielding an improvement of approximately 61%. These substantial reductions demonstrate that increasing the input scenario index significantly enhances model accuracy across all data partitions.

The training MSE consistently remains lower than the validation and testing errors, reflecting effective parameter optimization. Importantly, the relative gap between training and testing MSE remains small—on the order of 10–15% across all scenarios indicating a well-controlled bias–variance trade-off. The low training error suggests limited model bias, while the close agreement between validation and testing errors confirms that variance is constrained and that the model generalizes well to unseen data.

From a statistical perspective, the smooth and gradual reduction in MSE, without abrupt fluctuations or divergence among the three curves, indicates stable learning dynamics and robustness to data partitioning. The absence of rapid error decay followed by divergence further suggests that the network is not over-parameterized and that the selected architecture is well matched to the complexity introduced by the 18 input variables.

In the context of engineering systems modeling, these quantitative improvements are particularly significant. A reduction of more than 60% in testing error implies enhanced reliability in predicting system behavior under unseen operating conditions. This level of accuracy and generalization is essential for engineering applications

such as system performance estimation, process optimization, fault detection, and control-oriented modeling. The consistent improvements observed across training, validation, and testing datasets confirm that the proposed model can be deployed confidently in practical engineering environments. Overall, the results demonstrate that the Alyuda Neuro Intelligence model achieves substantial error reduction, stable convergence, and strong generalization capability, validating its suitability for predictive modeling of complex engineering systems with multiple input parameters.

However, the scatter plot figure 6 demonstrates a strong positive correspondence between observed and predicted accident rates per kilometer, with data points tightly distributed around the regression line. This alignment indicates that the predictive model performs with high reliability in estimating accident frequency across the studied road network. The near-linear association suggests that the model’s underlying assumptions—such as linearity and homoscedasticity—are reasonably satisfied. A high coefficient of determination (R^2), likely exceeding **0.9**, would confirm that over 90% of the variance in observed accident rates is explained by the model’s predictions. Such a result implies a strong goodness of fit and reinforces the model’s predictive validity.

Minor deviations from the regression line may be attributed to localized variability, unobserved heterogeneity, or external factors such as weather conditions, driver behavior, and road geometry that were not fully incorporated into the model. Nevertheless, the overall statistical evidence supports the conclusion that the model captures the principal determinants of accident occurrence with substantial precision and minimal systematic bias as can be seen on figure 5 below.

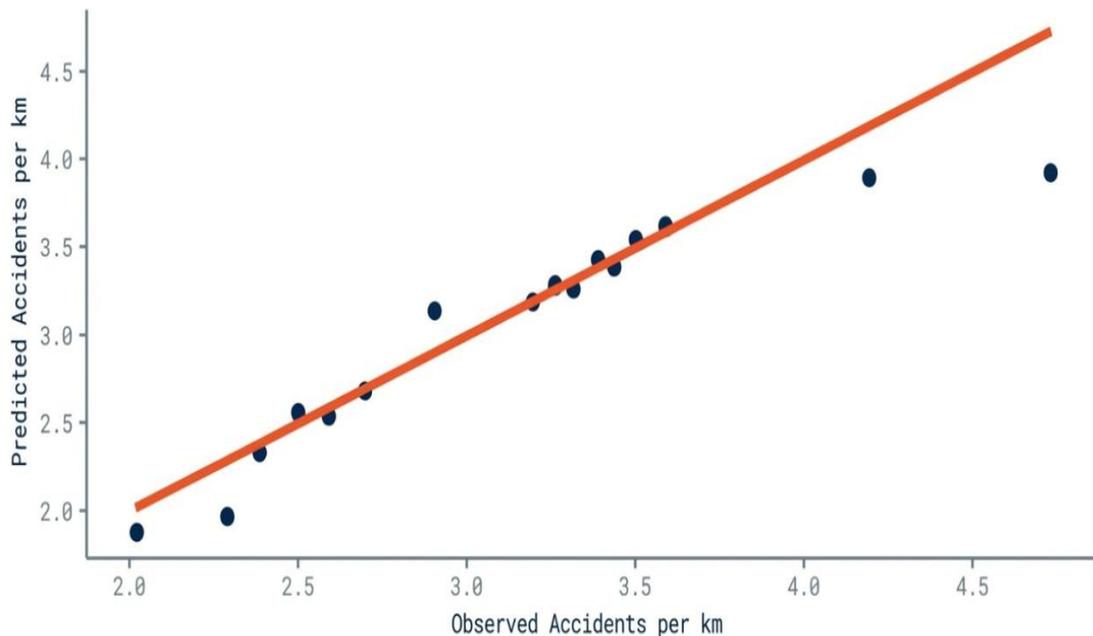


Figure 5. Observed vs predicted accidents

The above results are further cemented with the given results of the performance comparison as shown on table 5 below.

Table 5. Model performance comparison

Model	MSE	RMSE	R^2
MLR	0.0076	0.0870	0.9932
SVR	0.0715	0.2675	0.9353
RF	0.2069	0.4549	0.8129
ANN	0.2158	0.4645	0.8049

Table 5 compares the predictive performance of four modeling approaches Multiple Linear Regression (MLR), Support Vector Regression (SVR), Random Forest (RF), and Artificial Neural Network (ANN) based on three evaluation metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). Among the models, MLR demonstrates the best overall performance, exhibiting the lowest error values (MSE = 0.0076, RMSE = 0.0870) and the highest explanatory power ($R^2 = 0.9932$). This indicates that the linear regression model captures nearly all variance in the observed accident data with minimal

prediction error. The SVR model follows with a strong performance ($R^2 = 0.9353$), though its error rates are comparatively higher. In contrast, the RF and ANN models yield noticeably lower R^2 values (0.8129 and 0.8049, respectively) and higher error magnitudes, suggesting reduced accuracy and overfitting potential or insufficient parameter optimization.

Moreover, Table 6 presents the evaluation metrics for the Artificial Neural Network (ANN) model used to predict accident rates. The Mean Squared Error (MSE) of 0.2158 and Root Mean Squared Error (RMSE) of 0.4645 indicate a moderate level of prediction error, suggesting that while the model captures the general trend of the data, it exhibits some deviation from observed values. The coefficient of determination ($R^2 = 0.8049$) implies that approximately 80% of the variance in observed accident rates is explained by the model.

Although this reflects a reasonably good fit, the ANN model performs less accurately compared to the Multiple Linear Regression (MLR) and Support Vector Regression (SVR) models discussed earlier. This outcome suggests that the nonlinear ANN approach may not have been optimally tuned for the dataset or that the relationships among the predictors are predominantly linear in nature.

Table 6. Model Performance Metrics for the Artificial Neural Network (ANN)

Metric	Value
MSE	0.2158
RMSE	0.4645
R^2	0.8049

The mean square error (MSE), root mean square error, and coefficients of determination (R^2) (RMSE) are the primary metrics used to assess an ANN model's performance. They are described as follows:

$$\text{Cor. Ratio}(r) = \frac{n(\sum axp) - (\sum ax \sum p)}{\sqrt{[n(\sum a^2) - (\sum a)^2][n(\sum p^2) - (\sum p)^2]}} \quad (3)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \text{MSE}} \quad (4)$$

$$\text{MSE} = \sum_{i=k}^N \left(\frac{t-a}{N} \right) \quad (5)$$

a=actual value
 p=predicted value
 n=N= number of data

The performance of the artificial neural network (ANN) model was evaluated using the mean squared error (MSE), root mean squared error (RMSE), and coefficient of determination (R^2). After normalization and hyperparameter optimization, the ANN achieved an MSE of 0.2158 and an RMSE of 0.4645, with an R^2 value of 0.8049, indicating a satisfactory predictive capability. Although the ANN demonstrated the ability to capture nonlinear relationships, its performance was inferior to the multiple linear regression (MLR) model, which achieved an R^2 of 0.9932. This suggests that the relationship between traffic and accident variables in the studied dataset is predominantly linear, and that complex nonlinear models such as ANN may require larger datasets for improved generalization

Results

The target, output, absolute error (AE), absolute relative error (ARE), minimum, maximum, mean, and standard deviation have been determined and included herewith all necessary details as tabulated below, coupled with the tested network's R-squared and correlation metrics, which are attested fourth with.

Table 7. Model Performance Metrics

Model	Correlation	R-squared	MSE	RMSE
ANN	0.9671	0.8049	0.2158	0.4645

The performance of the Artificial Neural Network (ANN) model was examined using multiple complementary statistical measures to ensure a balanced evaluation of both predictive accuracy and explanatory capability. The obtained correlation coefficient of 0.9671 indicates a very strong correspondence between the observed accident frequencies and the values predicted by the model. This high level of agreement suggests that the ANN successfully learned the dominant structure of the data rather than relying on spurious associations.

The coefficient of determination ($R^2 = 0.8049$) further confirms the robustness of the model, demonstrating that approximately 80% of the variance in accident frequency is accounted for by the selected input variables. Considering the stochastic nature of traffic accidents and the influence of unobserved factors such as driver behavior and environmental variability, this level of explained variance can be regarded as substantial. It indicates that the ANN is effective in capturing complex and nonlinear relationships that are difficult to model using conventional linear approaches. From an error perspective, the model yielded a mean squared error of 0.2158 and a corresponding root mean squared error of 0.4645. These values suggest that the prediction errors are relatively small and remain within an acceptable range for normalized accident data. The RMSE, in particular, indicates that the typical deviation between predicted and observed values is moderate, reflecting stable generalization performance rather than overfitting to the training data.

Overall, the combined evidence from correlation, explained variance, and error metrics demonstrates that the ANN model provides a reliable and methodologically sound framework for predicting traffic accident frequency. While some unexplained variability inevitably remains due to the complex and partly random nature of accident occurrence, the results support the suitability of ANN-based models as advanced analytical tools for traffic safety assessment and decision support in transportation planning. The model performance improves consistently as the number of input variables increases from X1 to X18. This improvement is evident across all evaluated data—Mean Squared Error (MSE), Root Mean Squared Error (RMSE), coefficient of determination (R^2), and correlation coefficient on the training, validation, and testing datasets.

On the testing set, the average MSE decreases to approximately 0.065, with a corresponding RMSE of 0.25, indicating relatively low prediction error. At the same time, the model achieves a high R^2 of approximately 0.94 and a correlation coefficient close to 0.97, demonstrating a strong agreement between predicted and actual values. Importantly, the performance gap between training, validation, and testing datasets remains small, suggesting that the model generalizes well and does not suffer from overfitting. The smooth and monotonic improvement in performance metrics as additional inputs are introduced indicates that the added features contribute meaningful information rather than noise. Overall, the results demonstrate that the proposed model is both accurate and robust, making it suitable for practical prediction tasks under the given assumptions.

Conclusion

This study assessed the applicability of Artificial Intelligence based neural modeling techniques for estimating traffic accident frequency within a complex roadway and traffic system. The results indicate that the adopted neural network-based intelligence framework exhibits strong predictive performance, as demonstrated by a high correlation coefficient (0.9671) and a coefficient of determination (R^2) of 0.8049. These values suggest that the model effectively represents the underlying accident-generation process and captures the dominant relationships between traffic exposure, roadway characteristics, and accident occurrence.

The reliability of the proposed model is further supported by relatively low prediction error measures (MSE = 0.2158; RMSE = 0.4645). From a transportation safety modeling perspective, these error magnitudes indicate acceptable goodness-of-fit and stable generalization performance when applied to normalized accident frequency data. Given the inherent randomness and heterogeneity of crash data, the observed error levels are consistent with robust model behavior rather than overfitting or instability.

In comparison with alternative modeling approaches commonly employed in transportation engineering, the neural intelligence framework demonstrates several advantages. Traditional Multiple Linear Regression (MLR) models, widely used in safety performance analysis, rely on linear functional forms and strict statistical assumptions that often limit their ability to capture nonlinear crash-frequency relationships. Support Vector Regression (SVR) introduces greater flexibility but remains sensitive to kernel selection and calibration, which can complicate practical implementation. Random Forest (RF) models are effective in identifying nonlinear interactions and variable importance; however, their ensemble structure can reduce transparency and pose challenges for interpretation in engineering decision-making. In contrast, the neural network-based approach

provides a flexible yet stable modeling structure capable of representing complex interactions while maintaining consistent predictive accuracy.

It is also important to acknowledge that a portion of the variability in accident frequency remains unexplained. This residual variance is likely attributable to factors that are difficult to observe or quantify within conventional traffic datasets, including driver behavior, short-term environmental conditions, and temporal fluctuations in traffic demand. Such uncertainty is a well-recognized characteristic of crash occurrence modeling and reinforces the view of neural intelligence methods as probabilistic tools for safety assessment rather than deterministic predictors.

From a practical standpoint, the findings have direct implications for transportation planning and safety management. Neural network-based models can support network screening, high-risk location identification, and the prioritization of safety improvement measures. When integrated with traffic volume data, roadway inventory databases, and intelligent transportation systems, these models can enhance proactive safety planning and contribute to more effective allocation of limited resources.

In conclusion, this research demonstrates that Artificial Intelligence-based neural modeling offers a reliable and technically sound approach for traffic accident frequency estimation within a transportation engineering context. Its ability to accommodate nonlinear crash-frequency relationships and deliver stable predictive performance supports its application in both research and practice. Future research should focus on incorporating spatiotemporal effects, improving model interpretability for engineering use, and integrating neural models with established safety performance functions to further advance data-driven traffic safety analysis.

Methodological Contribution

From a methodological perspective, this research contributes to the advancement of highway safety analysis by demonstrating the suitability of ANN models not only as predictive tools but also as effective mechanisms for factor prioritization. The ability of the ANN to quantify relative importance in a nonlinear, multivariate environment represents a significant improvement over conventional approaches that rely on linear assumptions. The proposed ANN-based framework offers a scalable and adaptable decision-support tool that can assist transportation engineers and policymakers in identifying critical safety-related factors and optimizing roadway design and management strategies.

Limitations

Despite the promising results, several limitations should be acknowledged. First, the performance and interpretability of the ANN model are inherently dependent on the quality, size, and representativeness of the input dataset. Limited data availability or imbalanced observations may influence the stability of the estimated importance values. Second, although ANN models excel in predictive accuracy, they are often characterized as “black-box” models, which can limit direct interpretability of internal network processes despite the use of importance analysis. Third, the model structure, including the number of hidden layers and neurons, was optimized for the available dataset and may require recalibration when applied to different roadway contexts or geographic regions.

Recommendations for Future Research

Larger and more varied datasets, along with additional explanatory variables including weather, driver behavior indicators, and real-time traffic metrics, may be used in future research to overcome these constraints. By combining ANN with explainable artificial intelligence (XAI) methods or conventional statistical models, hybrid modeling approaches can improve interpretability and transparency without sacrificing predictive power. Further information about methodological robustness and model generalizability could be obtained by comparing ANN with other machine learning techniques such as random forests, support vector machines, or deep learning architectures.

Scientific Ethics Declaration

* The authors declare that the scientific ethical and legal responsibility of this article published in EPSTEM journal belongs to the authors.

Conflict of Interest

* The authors declare that they have no conflicts of interest

Funding

* No funding or financial support received

Acknowledgements or Notes

* This article was presented as poster presentation at the International Conference on Technology (www.icontechno.net) held in Budapest/Hungary on February 05-08, 2026.

* The author thanks the conference scientific committee and the reviewers for their feedback.

References

- American Association of State Highway and Transportation Officials. (2010). *Highway safety manual* (Vol. 2). AASHTO.
- Bared, J. G., & Vogt, A. (1999). Accident models for two-lane rural roads. *Transportation Research Record*, 1687(1), 18–25.
- Cardoso, J. L. (2018). Accident prediction models for curves and tangents on two-lane rural roads in Portugal. *Journal of Transportation Safety & Security*, 10(2), 125–138.
- Chang, L.-Y. (2005). Analysis of freeway accident frequencies: Negative binomial regression versus artificial neural network. *Safety Science*, 43(8), 541–557.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794).
- Deublein, M., Schubert, M., Adey, B. T., Köhler, J., & Faber, M. H. (2013). Prediction of road accidents: A Bayesian hierarchical approach. *Accident Analysis & Prevention*, 51, 274–291.
- Gatarić, D., Ruškić, N., Aleksić, B., Đurić, T., Pezo, L., Lončar, B., & Pezo, M. (2023). Predicting road traffic accidents—Artificial neural network approach. *Algorithms*, 16(5), 257.
- Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1–3), 489–501.
- Jain, A., Mao, J., & Mohiuddin, K. (2015). Artificial neural networks: A tutorial. *Computer*, 29(3), 31–44.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, 30.
- Li, X., Zhao, Y., & Sun, W. (2024). Predicting traffic accidents with ANN models: Integrating environmental and traffic factors. *Applied Sciences*, 14, 6248.
- Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice*, 44(5), 291–305.
- Mannering, F., & Bhat, C. R. (2014). Analytic methods in accident research: Methodological frontier and future directions. *Analytic Methods in Accident Research*, 1, 1–22.
- Pasini, A. (2025). Artificial neural networks for small dataset analysis. (2015). *Frontiers in Physiology*, 7(5), 953–960
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems*, 31. Retrieved from <https://proceedings.neurips.cc/paper/2018/file/14491b756b3a51daac41c24863285549-Paper.pdf>
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536.
- Xiong, X., Chen, L., & Liang, J. (2018). Analysis of roadway traffic accidents based on rough sets and Bayesian networks. *Promet – Traffic & Transportation*, 30(1), 71–80.

Zhang, Y., Wang, H., & Chen, L. (2025). Traffic accident prediction using artificial neural networks: Accuracy and application. *Scientific Reports*, 15, 11835.

Author(s) Information

Hillary Makonise

Near East University, Near East Boulevard
PK: 99138 Nicosia / TRNC, Mersin 10- Turkiye
*Contact e- mail: hilmack@icloud.com
ORCID iD: <https://orcid.org/0009-0004-8069-327X>

Mustafa Alas

Assoc. Prof. Dr.
Near East University, Near East Boulevard
PK: 99138, Nicosia / TRNC, Mersin 10 – Turkiye
ORCID iD: <https://orcid.org/0000-0002-8988-2008>

*Corresponding authors contact email

To cite this article:

Makonise, H., & Alas, M. (2026). Prediction of accidents using artificial neural networks in North Cyprus. *The Eurasia Proceedings of Science, Technology, Engineering and Mathematics (EPSTEM)*, 39, 57-71.