Jordan Journal of Civil Engineering, Volume **20**, No. **2**, **2026**
DOI: https://doi.org/10.14525/JJCE.v20i2.06

Jordan Journal of Civil Engineering

Journal homepage: https://jjce.just.edu.jo

Analysis of Pedestrian Accidents Using Machine Learning: Jordan as a Case Study

Laith D. AlOmari [1], ***Mohammed Taleb Obaidat*** [2]*

[1] *Graduate Student, Civil Engineering Department, Engineering College, Jordan University of Science and Technology (JUST); P.O. Box 3030, Irbid, Jordan. E-Mail: ldalomari23@eng.just.edu.jo*
[2] *Prof. of Civil Engineering, Civil Engineering Department, Engineering College, Jordan University of Science and Technology (JUST); P.O. Box 3030, Irbid, Jordan. * Corresponding Author. E-Mail: mobaidat@just.edu.jo*

ARTICLE INFO

Article History:
Received: 14/7/2025
Accepted: 17/10/2025

ABSTRACT

Jordan continues to face a serious public health challenge due to pedestrian-related traffic accidents, which contribute significantly to human and economic losses. This study aims to investigate the historical characteristics of pedestrian accidents in Jordan and to develop predictive models using both classical regression techniques and artificial neural networks (ANNs). Pedestrian accident data was collected from official annual reports published by the Jordan Traffic Institute (JTI), spanning the years from 1984 to 2022. Descriptive analysis focused on the period from 2013 to 2022, while model development utilized data from 1984 to 2022.

The analysis revealed that although the fatality risk has decreased over the years from 13 to 6 deaths per 100,000 populations between 2013 and 2022, it remains relatively high compared to global standards. Children aged 3-5 years and elderly individuals over 60 years of age were identified as the most vulnerable age groups. Pedestrian accidents were most frequent during peak evening hours and at low-speed limits of 40–60 km/h, often due to driver negligence or inadequate pedestrian infrastructure.

Three classical regression models were developed: linear, logarithmic, and power, with R^2 values of 0.937, 0.913, and 0.928, respectively. The linear model showed the best fit among the traditional approaches. Additionally, an ANN model with two hidden layers was trained using registered vehicle data as input, achieving an R^2 value of 0.981, indicating superior predictive performance and the ability to capture complex non-linear trends.

These findings highlight the critical role of advanced machine learning techniques in enhancing traffic safety planning and policy formulation. The study recommends integrating AI-driven models into national traffic monitoring systems and adopting urban planning strategies that prioritize pedestrian safety to reduce accidents and their severe consequences.

Keywords: Pedestrian accidents, Artificial intelligence, Machine learning, Regression, ANN, Safety, Jordan.

INTRODUCTION

Traffic accidents are multi-faceted events influenced by a combination of factors, including roadway design, environmental conditions, vehicle characteristics, and human behavior—both from drivers and pedestrians (Al-Masaeid, 2009). These incidents range in severity from minor injuries to catastrophic outcomes involving fatalities. According to the World Health Organization (WHO, 2018a), approximately 1.35 million people die each year due to road traffic accidents, while up to 50 million more sustain non-fatal injuries. The economic burden is also considerable, as traffic-related injuries cost most countries around 3% of their gross domestic product. Vulnerable road users, such as pedestrians, cyclists, and motorcyclists, account for more than a half of all road traffic fatalities, with 93% of these deaths occurring in low- and middle-income countries that collectively own only about 60% of the world's vehicles (WHO, 2018b).

Given this vulnerability, pedestrian safety should be a top priority in transportation policy and urban planning. However, many road networks in developing countries, including Jordan, are designed with a focus on vehicle throughput and speed, often overlooking pedestrian protection (O.Z.E.N., Sayin & Yuruk, 2017). This imbalance worsens the risk faced by pedestrians, particularly in regions undergoing rapid urbanization and motorization.

In developing countries, injuries and fatalities from pedestrian-related traffic accidents are still a significant public health challenge and are projected to rise with continued population and vehicle growth (Sethi & Zwi, 1999). Jordan has experienced a sustained pattern of pedestrian-related incidents. Al-Omari and Obaidat (2013) analyzed pedestrian accidents in Irbid City between 1999 and 2001 and found that most incidents occurred in clear weather, on dry road surfaces, and outside intersections—often at relatively low-speed limits. Their analysis also revealed that children under 15 years of age and males were the most affected demographic groups.

With the rapid development of artificial intelligence (AI) and machine learning (ML), new avenues have appeared for analyzing traffic accident data more effectively. These technologies have revolutionized how large datasets are processed, offering predictive capabilities that far exceed traditional statistical methods (Hua, 2022). In the context of traffic safety, ML can identify hidden patterns in pedestrian accident data and support the development of more targeted safety strategies.

This study addresses the lack of research applying machine learning techniques to pedestrian accident analysis in Jordan. The main objectives of the paper are threefold: (1) to analyze the historical characteristics of pedestrian accidents in Jordan; (2) to develop and evaluate multiple regression models, including linear, logarithmic, power and artificial neural network (ANN) approaches to predict accident frequency; and (3) to provide actionable recommendations for enhancing pedestrian safety through data-informed planning and regulation. The analysis is based on accident data obtained from the Jordan Traffic Institute, covering the years from 2013 to 2022 for descriptive analysis and the years from 1984 to 2022 for model development.

LITERATURE REVIEW

Many studies have been conducted in this field, showing pedestrian accidents' main causes and characteristics. One of the early and notable studies by Fontaine and Gourlet (1997) proposed a typological framework for the in-depth analysis of pedestrian crashes. Their research identified several critical factors contributing to pedestrian accidents, including the pedestrian's age and sex, alcohol consumption, and the type of transportation mode involved. The study revealed that both children and the elderly are at higher risk and emphasized the importance of tailoring educational programs to specific demographic groups to mitigate pedestrian injuries.

Kouabenan and Guyot (2004) further examined pedestrian accident causality through an analysis of 55 accident reports, using three different analytical techniques to assess severity-based patterns. Their findings highlighted that those rural areas characterized by higher vehicle speeds pose significant risks to pedestrians. Additionally, limited visibility caused by environmental obstacles was found to be a common contributing factor. The researchers recommended expanding the sample size to enhance the generalizability and reliability of their conclusions.

Regionally, Al-Masaeid (2009) analyzed traffic accident trends in Jordan using annual reports from the Jordan Traffic Institute. His study combined both pedestrian and driver characteristics, finding children and the elderly as the most vulnerable age groups—findings consistent with global trends. This was later supported by Al-Omari, Ghuzlan and Hasan (2009), whose research showed that most pedestrian accidents in Jordan occur at speed limits below 50 km/h, with decreasing accident rates as speed increases.

In a more focused local context, Al-Omari and

Obaidat (2013) conducted a study analyzing pedestrian accidents in Irbid City, Jordan, using data collected between 1999 and 2001. Their work considered a wide range of variables, such as road surface conditions, lighting, weather, accident timing, and fault attribution. Results showed that children under 15 years of age accounted for the highest proportion of victims and that most accidents occurred during the noon-to-afternoon period on weekends and in the summer season—again, at low-speed limits.

More recently, the integration of machine learning (ML) in traffic safety studies has gained momentum. Pérez et al. (2023) investigated injury severity among elderly pedestrians in urban traffic accidents in Spain using ML algorithms. Their analysis compared logistic regression (LR) and random forest (RF) models. While RF demonstrated superior predictive performance, LR models offered better interpretability. Age was identified as a significant determinant of injury severity, supporting the call for more targeted studies on vulnerable demographic groups.

More advances in the application of machine learning (ML) and artificial intelligence (AI) to pedestrian safety analysis have enabled more sophisticated modeling of accident risk and injury severity. Ul Arifeen et al. (2023) utilized Bayesian neural networks (BNNs) alongside other ML algorithms, including conventional neural networks, random forests, and k-nearest neighbors, to predict the severity of vehicle-pedestrian crashes. Their research, focused on Hong Kong, found that BNNs outperformed other models in predictive accuracy, particularly in the presence of noisy or incomplete data. Notably, environmental conditions, such as weather and lighting, along with temporal variables, were identified as key contributors to crash outcomes. This supports the broader recognition that advanced ML models are well-suited for capturing the complex, non-linear relationships inherent in pedestrian accident data.

Similarly, Gálvez-Pérez et al. (2023) examined the severity of injuries sustained by elderly pedestrians in urban traffic accidents in Spain. By comparing logistic regression with random forest models, their study found that, although random forests provided a higher predictive power, logistic regression remained advantageous for its interpretability. The findings revealed that both accident-related factors (such as location and lighting) and built environment features significantly influenced injury outcomes. Interestingly, accidents occurring in larger cities tended to be less severe, while insufficient artificial lighting was linked to increased injury severity among elderly pedestrians. These results highlight the dual importance of model interpretability and predictive strength in pedestrian safety research.

Complementing these approaches, recent studies have used interpretable machine learning frameworks to visualize risk factors and spatial distributions. For example, Yoon (2025) applied XGBoost in conjunction with Local Interpretable Model-agnostic Explanation (LIME) to analyze large-scale crash data from Seoul. The study demonstrated how interpretable ML models can identify high-risk zones, uncover non-linear relationships, and provide actionable insights for urban policymakers through spatial heatmaps and factor importance plots. This aligns with the growing emphasis on using transparent AI methods to support targeted interventions and risk mitigation in pedestrian safety.

From a forensic perspective, Bastien et al. (2024) showcased the utility of machine learning in re-constructing pedestrian collisions where physical evidence is limited. By training ML models on thousands of simulated accident scenarios, they were able to accurately estimate collision parameters, such as impact speed and pedestrian gait even when traditional sources of evidence (e.g. video or event data recorders) were unavailable. Their tool, validated against real-world crash cases in the UK, achieved prediction errors under 10%, illustrating the potential of ML to support accident investigations and inform both legal proceedings and emergency response protocols.

Although these studies collectively provide valuable descriptive insights into pedestrian accident characteristics and demonstrate the potential of AI/ML tools, most of them remain context-specific and method-driven rather than integrative. Few attempts have synthesized cross-study findings to identify consistent risk factors, limitations in methodology, or gaps in generalizability across regions.

In particular, prior studies have often analyzed isolated variables (e.g. speed, age, weather) without embedding them in broader systemic frameworks that reflect the complexity of pedestrian safety. Moreover, the Middle Eastern context remains under-represented, with scarce research addressing local conditions, infrastructure characteristics, or socio-cultural patterns.

This creates a clear knowledge gap where modern AI-driven methods can be used, not only to predict accident severity, but also to inform data-driven planning and safety interventions tailored to the region.

Despite these contributions, relatively few studies have employed advanced AI or ML techniques to analyze pedestrian accident trends in the Middle East, and particularly in Jordan. This gap underscores the need for predictive modeling approaches that can help identify high-risk conditions and inform urban planning decisions. The present study responds to this gap by critically integrating prior findings, highlighting their methodological shortcomings, and positioning itself as a step toward developing a comprehensive, context-aware model for pedestrian accident analysis in Jordan.

METHODOLGY

The methodology adopted in this study was designed to ensure rigor, transparency, and reproducibility throughout all research phases. The process began with comprehensive data acquisition, where annual pedestrian accident records were collected from the official reports of the Jordan Traffic Institute (JTI) covering the years from 1984 to 2022. Two datasets were established for the analyses. The first dataset, used for descriptive analysis, covered the period from 2013 to 2022 and included detailed records of accident counts, fatality rates, injury severity, age groups, gender distribution, time of occurrence, speed zones, and related contributing factors. The second dataset, developed for predictive modeling, included annual records of pedestrian accidents and the number of registered vehicles, supplemented by supporting variables, such as annual population and total traffic accidents.

To ensure analytical integrity, the data underwent careful pre-processing and cleaning. Records containing missing or incomplete values in any relevant variable were excluded from the analysis. All data fields were standardized in terms of naming conventions and measurement units. The annual number of registered vehicles was selected as the primary independent variable (predictor), while the annual number of pedestrian accidents was defined as the dependent variable (target) for all predictive modeling tasks. Other supporting variables, such as population and fatalities, were used to provide context and validation for the models.

The analytical approach involved the development and evaluation of both classical regression models and machine learning techniques. Initially, three classical regression models; linear, logarithmic, and power, were formulated to characterize the relationship between registered vehicles and pedestrian accident counts. These models were chosen to capture a spectrum of possible trends, ranging from simple proportional increases to non-linear or diminishing effects. Model fitting was performed using the scipy.optimize.curve_fit function in Python, and the adequacy of each model was assessed using the coefficient of determination (R^2) and adjusted R^2, the latter correcting for model complexity and sample size. Diagnostic plots, including fitted regression curves and residual plots, were generated to visually inspect model fit and bias. In addition to R^2 and adjusted R^2, model fit was evaluated using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). These criteria are particularly suitable for comparing non-nested models and are recommended when only a single independent variable is available, as in this study. Models with lower AIC and BIC values were considered to have superior fit.

Recognizing the limitations of classical models in capturing complex, non-linear relationships, the study proceeded to develop a supervised artificial neural network (ANN) model using TensorFlow and Keras. The ANN was designed with a single input node corresponding to the standardized annual number of registered vehicles. The architecture consisted of two hidden layers comprising 16 and 8 neurons, respectively, each using the rectified linear unit (ReLU) activation function. The output layer, employing a linear activation function, generated predictions of the annual number of pedestrian accidents. The model was trained using the mean squared error (MSE) loss function, optimized with the Adam algorithm set to a learning rate of 0.0001. Early stopping was applied to prevent overfitting, and the dataset consisted of 39 annual records covering the period 1984–2022. Of these, 31 records (80%) were used for model calibration (training), while the remaining 8 records (20%) were reserved for validation (testing). This division was selected to ensure a sufficient sample size for training while preserving independent data for evaluating model generalizability. The final configuration of the ANN, including all hyper-parameters, was selected based on iterative tuning to achieve the best performance on the validation data.

To ensure consistent units, the number of registered vehicles, originally reported in thousands, was converted into absolute counts (×1000) before training. Both input and output variables were standardized to improve training stability and predictions were inverse transformed back to their original scale. The split was performed chronologically (80% training, 20% testing) without shuffling to preserve temporal integrity.

Model selection followed an ablation analysis comparing single-hidden-layer and two-hidden-layer architectures under identical training conditions. While a single hidden layer is theoretically sufficient, it yielded weaker convergence and lower validation R^2 in our dataset. In contrast, the two-hidden-layer configuration demonstrated higher predictive accuracy and more stable convergence, justifying its adoption.

Furthermore, to enhance interpretability, SHAP (KernelExplainer) and LIME were applied to the final ANN. These model-agnostic methods quantified and visualized the marginal influence of vehicle registrations on accident predictions, providing transparent explanations of the ANN's behavior. Figures presenting SHAP summary plots and LIME local explanations have been added to support this analysis.

For both the classical regression models and the ANN, model performance was evaluated using multiple statistical metrics. The primary metric was the coefficient of determination (R^2), which measures the proportion of variance explained by the model. Adjusted R^2 was also calculated to account for the number of predictors relative to sample size. For the ANN, additional evaluation metrics included the mean absolute error (MAE) and root mean square error (RMSE), both of which quantify prediction error. The training and validation loss curves of the ANN were carefully monitored throughout the training process to confirm proper convergence and to identify any signs of overfitting. All model equations and performance results were summarized in comparative tables to facilitate clear interpretation.

To further validate predictive performance, k-fold cross-validation was performed with $k = 3$, 5, and 7 for all regression models as well as the ANN. For each fold, the dataset was partitioned into k sub-sets, with k–1 folds used for training and one-fold used for testing. The process was repeated k times, and the average values of root mean square error (RMSE) and mean absolute error (MAE) were reported as additional performance metrics.

All data processing, statistical analysis, and model development tasks were carried out in Python 3.10.4. The data manipulation and cleaning procedures were implemented using the Pandas library, while NumPy and Scipy were used for statistical analysis and classical regression modeling. Visualization of results, including regression fits and residuals, was performed with Matplotlib and Seaborn. The ANN was developed, trained, and validated using TensorFlow and Keras. To enhance reproducibility, sample scripts and representative datasets are made available in the supplementary materials, enabling independent verification of all analytical steps and results.

This methodology was structured to provide a robust, interpretable, and reproducible framework for investigating pedestrian accident trends and building predictive models. However, certain limitations remain, notably the absence of high-resolution spatial data, which constrains the analysis to national-level, annual counts. Potential biases in the original reporting and the moderate sample size of the predictive dataset may also limit the generalizability of the findings. Despite these constraints, the adopted approach allows for meaningful comparison between traditional statistical models and modern machine learning techniques, contributing new insights into the field of traffic safety analysis.

It is important to note that the current modeling framework primarily relied on vehicle registrations as the main independent variable due to data availability constraints. While this approach provided valuable predictive insights, it does not fully capture the multi-factorial nature of pedestrian safety.

To further clarify the research design and ensure transparency in the analytical process, a methodological flowchart was developed to visualize the complete sequence of steps undertaken in this study. As shown in Figure 1, the workflow begins with data acquisition and pre-processing, proceeds through both classical regression modeling and artificial neural network (ANN) development, and concludes with comparative evaluation of model performance. This diagram provides a comprehensive overview of how raw data was systematically transformed into actionable analytical outputs, highlighting key stages, such as data cleaning, model training, validation, and the interpretation of results.

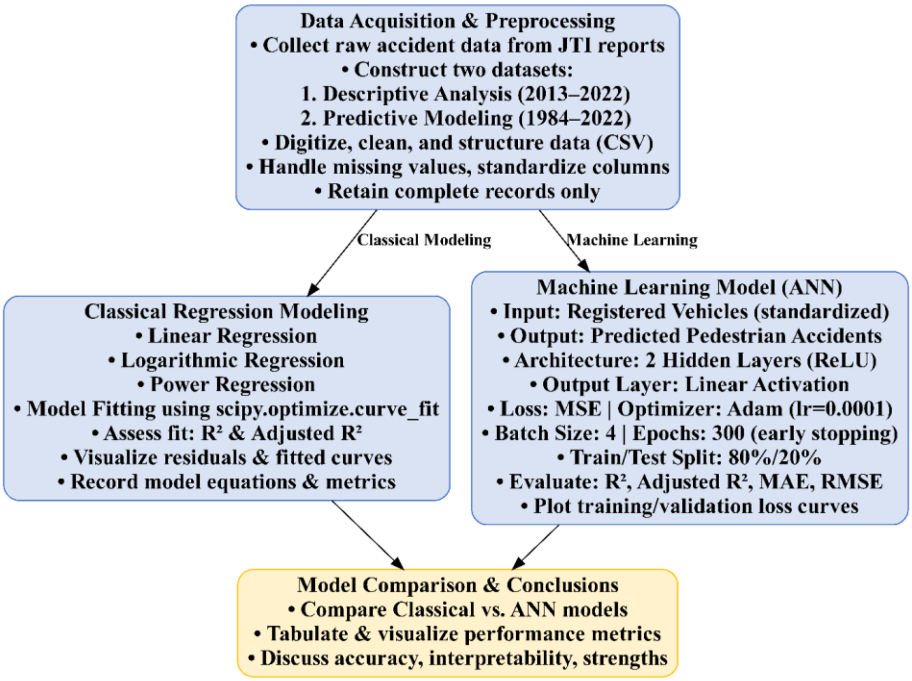

Figure 1. Methodological flowchart illustrating the data processing, modeling, and evaluation steps employed in this study

DATA COLLECTION

All data used in this research was collected from the Jordan Traffic Institute's (JTI) official annual reports covering the period from 1984 to 2022. The data was manually extracted, digitized, and structured to ease quantitative analysis and modeling. The historical data includes annual records of: Population (in thousands), Registered vehicles (in thousands), Total number of reported traffic accidents, and Fatalities resulting from traffic crashes. This information served as the foundation for both the descriptive and predictive components of the study. A sample of the data structure is presented in Table 1.

Table 1. Sample of annual traffic data from JTI reports (1984-1993)

Year	Population ($x10^3$)	Registered Vehicles ($x10^3$)	Total Accidents	Fatalities
1984	2595.1	211.657	15917	493
1985	2693.7	221.454	16078	527
1986	2796.1	232.361	14060	355
1987	2896.8	251.938	15884	396
1988	3001.1	249.527	18038	364
1989	3011.0	251.287	18336	355
1990	3453.0	254.617	17838	379
1991	3888.0	259.196	18756	379
1992	4012.0	276.301	20970	388
1993	4152.0	291.347	24799	440

ANALYSIS AND DISCUSSION

Jordan, located in the Middle East, has an area of 90 thousand square kilometers and a population of 11.3 million. It faces significant health problems due to traffic accidents. During the last ten years, total accidents increased from 10226 in 2013 to 11510 in 2022. However, for the same period, the number of pedestrian accidents varied between 2800 and 4000, showing an average percentage of 34.95% of total accidents. With this variation in accidents, there is an increase in population and the number of registered vehicles. Table 2 presents the growth of population, registered vehicles, total and pedestrian accidents between 2013 and 2022.

Table 2. Inhabitants, registered vehicles, total human injury and pedestrian accidents in Jordan

Year	Inhabitants (thousands)	Registered Vehicles	No. of Total Accidents	No. of Pedestrian Accidents
2013	6530	1263754	10226	3954
2014	6675	1331563	9756	3755
2015	9531	1412817	9712	2832
2016	9798	1502420	10835	3682
2017	10053	1583458	10446	3633
2018	10309	1637981	10431	3516
2019	10554	1677061	10857	3661
2020	10806	1729343	8451	2881
2021	11057	1795215	11241	3703
2022	11310	1855901	11510	3847

As shown in this table, there is a continuous increase in traffic accidents in population and the number of registered vehicles. However, pedestrian accidents vary in the number of these accidents, mostly around four thousand. Figure 2 demonstrates a continuous reduction in the fatality risk for Jordan and many other developed and developing countries. According to data of 2013, the fatality rate in Jordan was 13 compared with 31.5 deaths per 100,000 habitants in Oman and 1.8 in the UK. It reduced to 6 in 2022, showing that Jordan's strategies in preventing fatal accidents are practical, but still face a severe accident problem due to the high fatality rate.

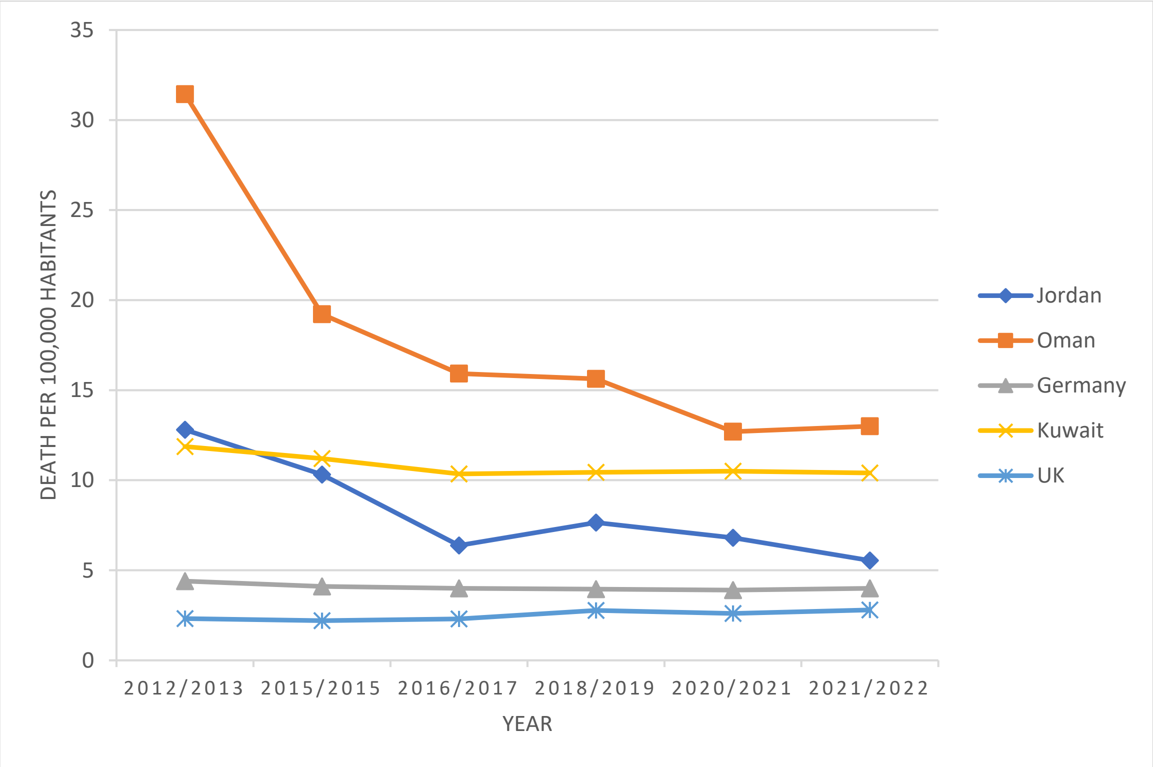

Figure 2. Death risks for a group of countries

Table 3 shows the number of pedestrian accidents and their casualties in Jordan from 2020 to 2022. Comparing the number of pedestrian accidents to their casualties shows that they resulted in a vast number of injuries and fatalities, which is expected, because road traffic regulations are looking for faster and uninterrupted traffic flow, ignoring the importance of pedestrian safety (O.Z.E.N. et al., 2017). Moreover, this can be related to inadequate urban planning, with towns developed along significant roads or schools located on main roads, causing increased hazardous situations (Al-Masaeid, 2009). In 2020, the table showed a reduction in accidents and their casualties; this can be explained by the restriction periods that occurred during the COVID-19 pandemic; according to da Silva et al. (2022), there is a significant reduction in the number of accidents observed during COVID-19 pandemic period compared with pre-pandemic period (da Silva et al., 2022).

Table 3. Pedestrian accidents and their casualties in Jordan (2020-2022)

Year	No. of Pedestrian Accidents	Pedestrian Casualties			
		Fatal	Serious injury	Medium injury	Slight injury
2020	2881	164	224	1594	1541
2021	3703	208	279	2017	1909
2022	3847	203	332	2167	1785

Figure 3 shows the involvement rate for pedestrians in cases of being killed or having a severe injury during the years 2013-2022. The involvement rate was determined as the number of fatal or seriously injured people in each age group relative to the population of the same age group. It is indicated that the child age group from 2 to 5 years old and the elderly age group over 60 years are the riskiest pedestrian age groups. Compared with the results of (Al-Masaeid, 2009), the results are almost the same, showing that those age groups need a quick and effective solution to reduce their risk rate through training and education programs for children and allocation of pathways for elderly age groups in order to reduce their interaction with the road due to their health conditions.

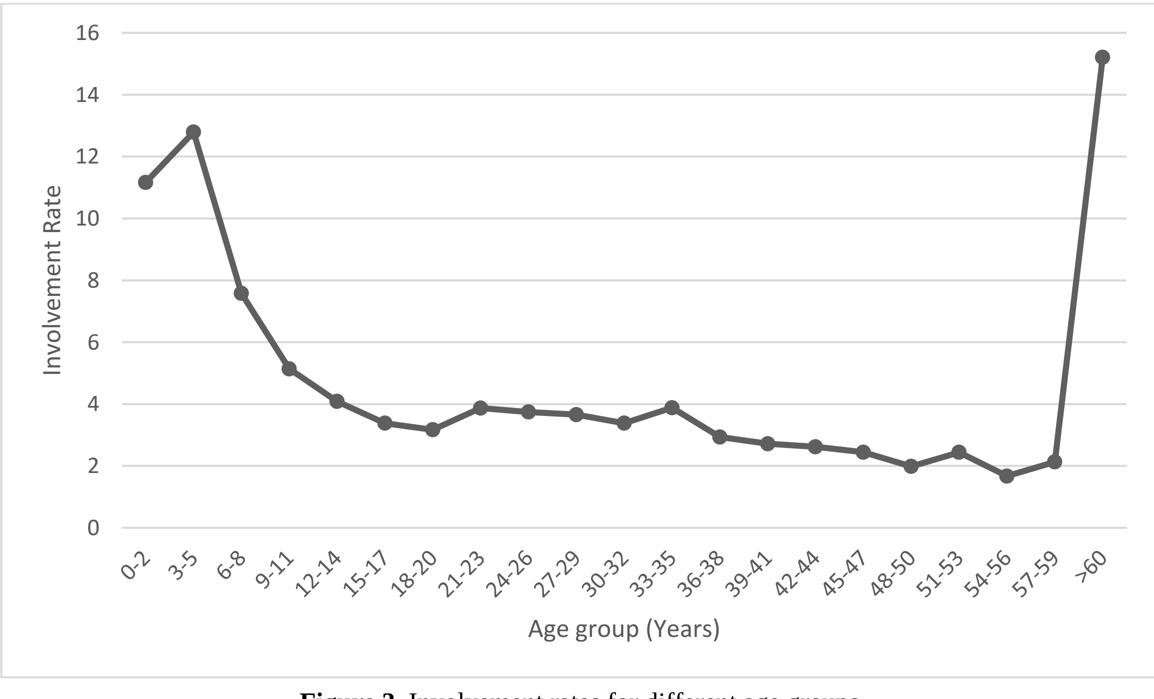

Figure 3. Involvement rates for different age groups

Figure 4 shows the involvement rate for pedestrians according to today's times. It was from indicated that the most hazardous time interval was from 4:00 PM to 11:00 PM, then the involvement rate started to decrease until getting the minimum value at 7:00 AM and started to increase again, which shows a concurrence with the (Al-Omari & Obaidat, 2013) result. This trend can be explained as 4:00 PM is considered the peak-hour period in most of Jordan's cities, which causes heavy traffic on the road. It is known that traffic density increases, and the possibility of accident occurrence increases too. Then, during the night, some of the highway's lighting is insufficient, causing a reduction in sight distance, especially on curves, which increases the possibility of pedestrian hitting. If the speed is high, the probability of fatal accidents increases. To solve this problem, it is suggested to apply a separation strategy for pedestrian and traffic movements for heavy traffic situations and provide sufficient lighting during the night, especially on major roads.

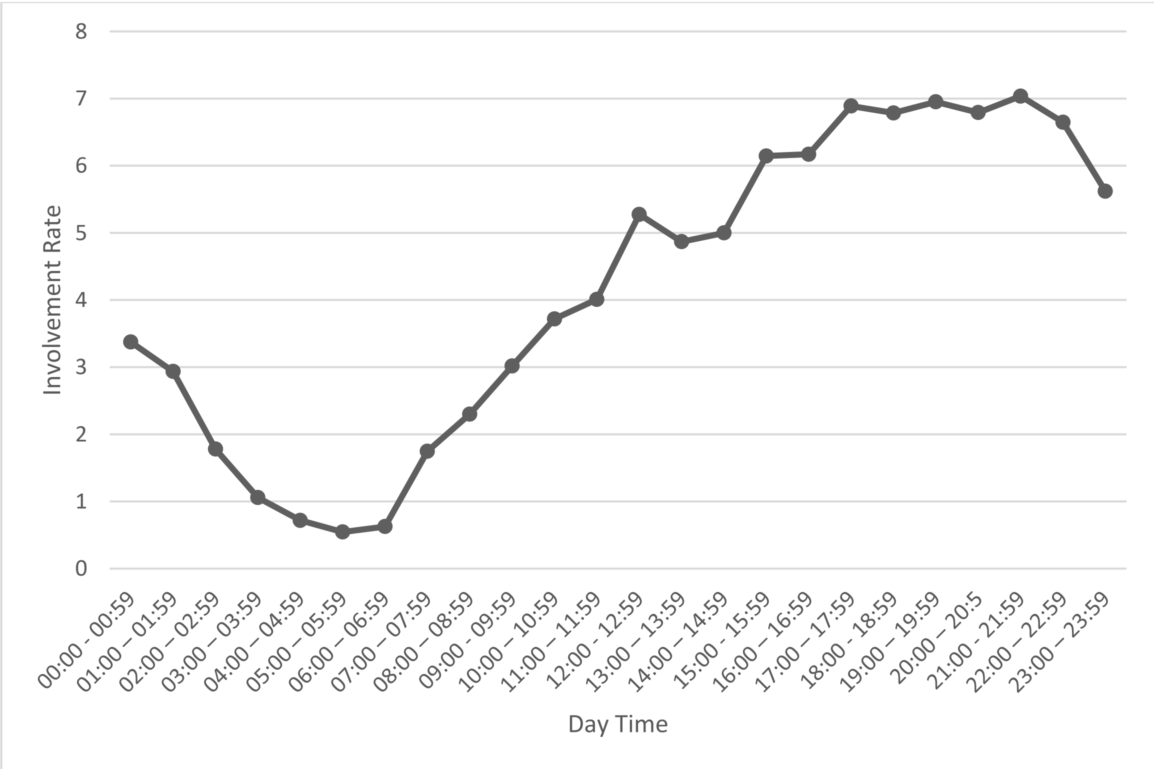

Figure 4. Involvement rates for different day times

Figure 5 shows the pedestrian involvement rate for different speed limits. The figure indicates that most pedestrian accidents occur at low-speed limits of 40-60 km/hour. However, drivers disregard priorities at low speeds, causing the probability of pedestrian hitting to occur. According to Koushki et al. (2001), a study conducted in Kuwait showed that higher speeds cause 17% of pedestrian accidents, while speed was considered the first cause of fatal accidents in pedestrians (Koushki et al., 2001). Also, according to Knowles et al. (2012), most of the fatalities in pedestrians are due to pedestrian collisions in London, where around 90% occurred at a speed limit of less than or equal to 48 km/hour (Knowles et al., 2012).

Table 4 shows different pedestrian actions that are considered as causes of pedestrian accidents. The action that is considered the primary cause of pedestrian accidents is walking on the road, contributing to 85.6% of all accidents. Walking on a road without sidewalks falls into the second place, followed by walking on the sidewalk, working on the road, walking in the opposite direction of traffic, and lastly, walking or standing on the median.

Table 4. Causes of pedestrian accidents (2013-2022)

Cause	Percentage
Walking on the road	85.6%
Walking on the road without sidewalks	4.7%
Walking on the sidewalk	3.2%
Working on the road	2.9%
Walking on the road on the opposite direction of traffic	1.4%
Walking or standing on the median island	0.7%
Others	1.4%

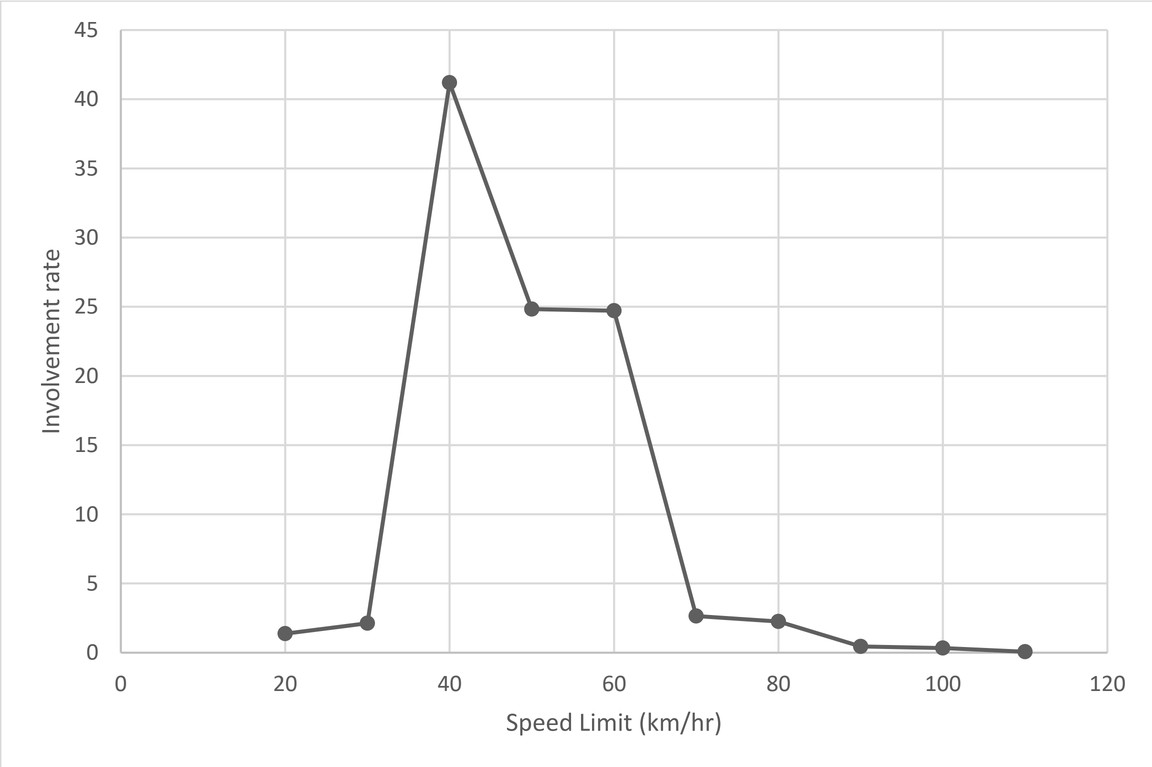

Figure 5. Involvement rates for different speed limits

Three classical regression models; linear, logarithmic, and power, were developed to assess the relationship between the number of registered vehicles and the number of accidents. The evaluation criteria included the coefficient of determination (R^2) and the adjusted R^2, which accounts for the number of predictors and sample size.

The linear model showed a strong fit with an R^2 value of approximately 0.937, indicating that around 93.7% of the variability in pedestrian accidents can be explained linearly by the increase in registered vehicles. This reflects a direct and proportional relationship.

The logarithmic model performed slightly worse than the linear one, with an R^2 value of 0.913. This suggests diminishing marginal increases in accidents with increasing vehicle registrations, which could imply that accident rates may stabilize after a certain vehicle threshold.

The power model, although theoretically appealing in transportation studies, achieved an R^2 of 0.928, which is very close to the linear model. The power form might reflect an underlying non-linear growth pattern, but is statistically less favored compared to the linear model in this dataset.

The adjusted R^2 values for all three models followed the same ranking, with the linear model retaining the highest value. This confirms the robustness of the linear relationship in this specific scenario, where the number of accidents increases nearly proportionally to the number of registered vehicles in Jordan over time.

In addition to classical models, an Artificial Neural Network (ANN) was trained using TensorFlow to explore a more flexible, data-driven approach. The ANN architecture consisted of one input node representing the scaled number of registered vehicles, two hidden layers (with 16 and 8 neurons, respectively) using ReLU activation functions, and one output node predicting the number of accidents, using a linear activation function. The network was trained with MSE loss and the Adam optimizer. Figure 6 illustrates the training and validation loss during the ANN training phase, confirming effective convergence and model stability.

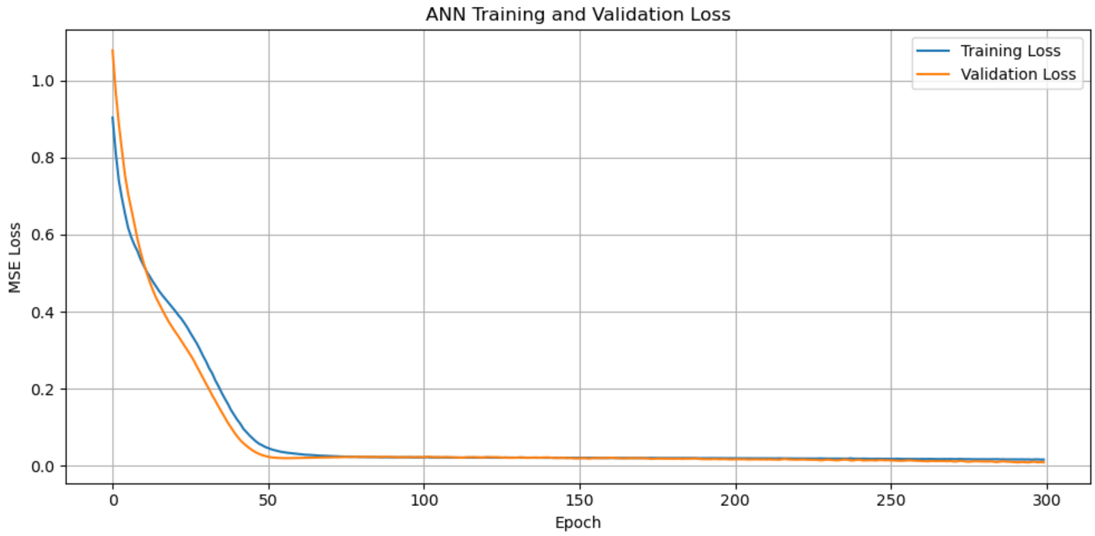

Figure 6. ANN training and validation losses

The ANN achieved an R^2 value exceeding 0.98, clearly outperforming all classical models. This significant improvement can be attributed to the model's ability to learn complex non-linear relationships that traditional regressions may miss. Moreover, the training *vs.* validation loss curve demonstrated excellent convergence with no sign of overfitting, further confirming the reliability of the network.

In addition to R^2 and adjusted R^2, model comparison was extended using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), as recommended when a single independent variable is involved. As shown in Table 5, the logarithmic model achieved the lowest AIC (844.8) and BIC (848.1) values, outperforming both the linear (AIC = 865.9, BIC = 869.3) and power (AIC = 880.8, BIC = 884.1) forms. This indicates that pedestrian accident growth follows a diminishing marginal trend with respect to vehicle registrations. Nevertheless, the ANN model clearly outperformed all classical regressions in predictive accuracy (R^2 = 0.981), highlighting the value of AI-based approaches for traffic safety analysis in Jordan. Table 5 summarizes the equations and performance metrics of the classical and ANN regression models used in this study, where y is the predicted number of accidents and x is the number of registered vehicles.

Table 5. Equations and performance metrics of the classical and ANN regression models

Model	Equation	R^2	Adjusted R^2	AIC	BIC
Linear	$y = 3.7960\times10^{-2} * x + 1.4710\times10^{3}$	0.937	0.935	865.9	869.3
Logarithmic	$y = 6.7590\times10^{3} * \ln(x) - 4.6494\times10^{4}$	0.913	0.910	844.8	848.1
Power	$y = 2.2272\times10^{-2} *$ x^1.1671	0.928	0.926	880.8	884.1
Artificial Neural Network (ANN)	Predicted by ANN (2 hidden layers with ReLU activations)	0.981	0.980	---	---

To verify the normality of regression residuals, Q–Q plots were generated using standardized residuals for all models, as illustrated in Figure 7. The plots show that the residuals of the linear, logarithmic, and power models are symmetrically distributed around the 45° reference line, confirming approximate normality. Although slight deviations are observed in the tails, the overall alignment supports the validity of the regression assumptions and confirms the reliability of the fitted models.

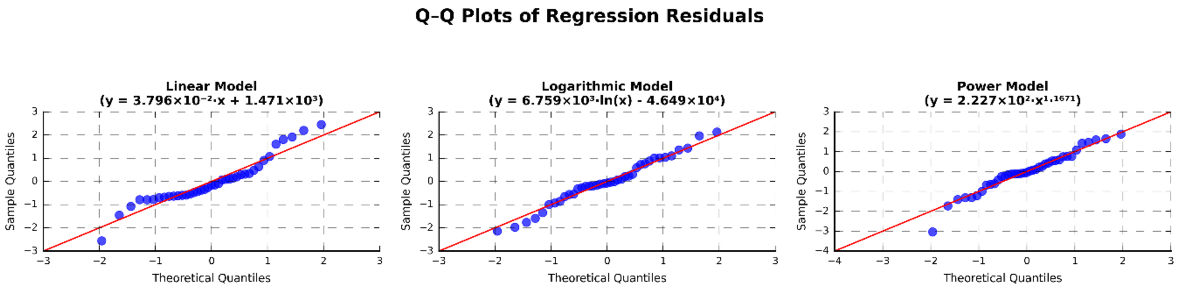

Figure 7. Q–Q plots of regression residuals for linear, logarithmic, and power models

An ablation study comparing different architectures confirmed that the two-hidden-layer network consistently outperformed single-hidden-layer alternatives, offering better validation R^2 and more stable convergence, thus validating its selection.

In addition, interpretability analyses were introduced to address the reviewer's concerns. SHAP and LIME were applied to the ANN predictions. The SHAP summary plots showed a strong, monotonic positive effect of vehicle registrations on accident counts, while LIME explanations around test samples confirmed locally linear and coherent surrogates. These findings enhance transparency and demonstrate that the ANN's predictions are not only accurate, but also explainable.

SHAP analysis shown in Figure 8 illustrates a monotonic positive effect of vehicle registrations on predicted accident counts, with all test samples contributing positively relative to the background mean. Complementary LIME local explanations in Figure 9 yielded positive slopes and high local R^2 values, confirming that in the vicinity of test instances, increased vehicle registrations lead to higher predicted accident counts.

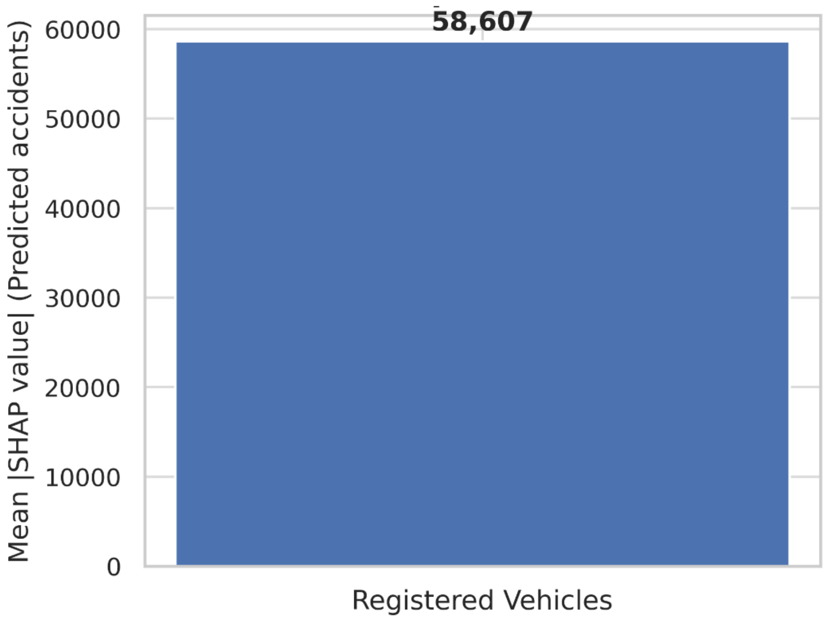

Figure 8. SHAP feature importance for registered vehicles (single feature)

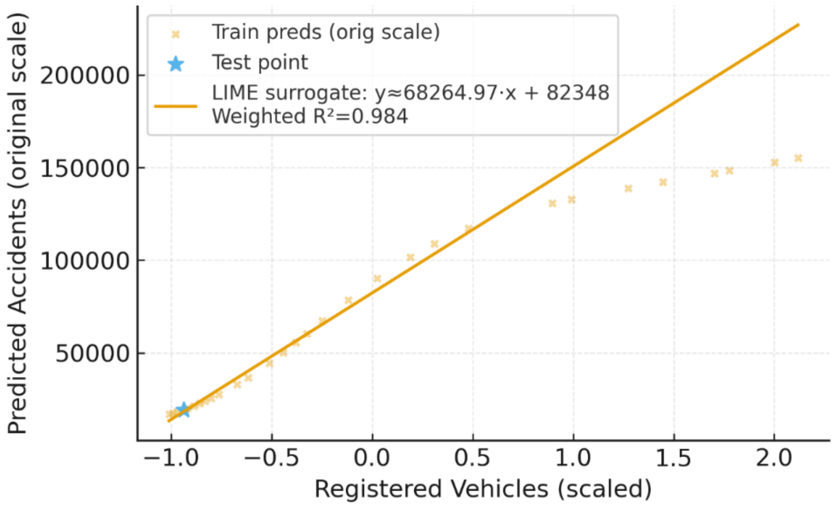

Figure 9. LIME local explanation (test instance)

Figure 10 shows the architecture of the ANN used in this study, including the input layer, hidden layers with ReLU activation, and the output layer with linear activation.

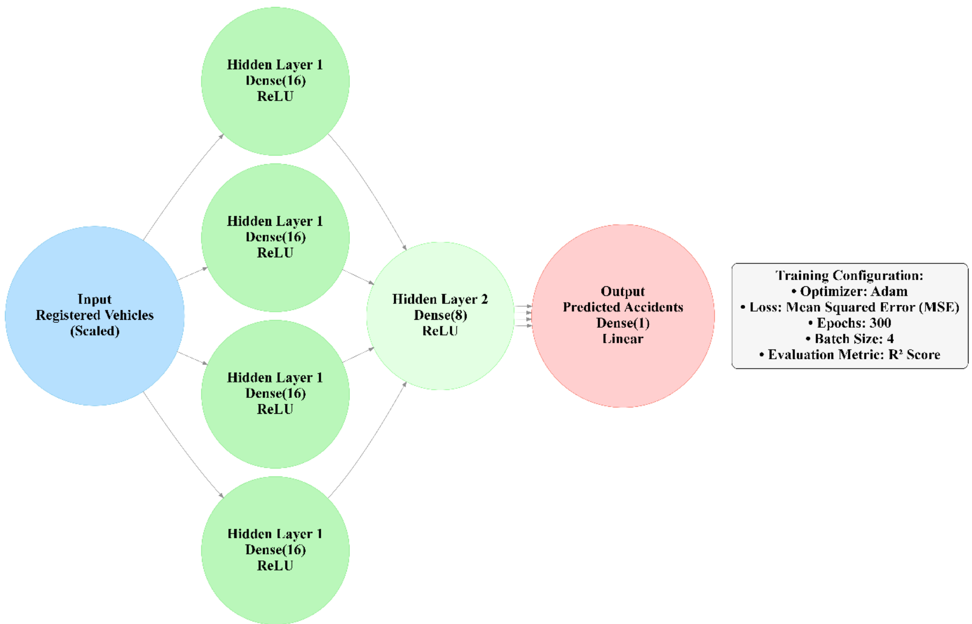

Figure 10. The architecture of the ANN

To further validate the predictive performance of the models, k-fold cross-validation (k = 3, 5, and 7) was conducted. Table 6 presents the RMSE and MAE values across folds. The results confirm that the logarithmic model achieved the lowest average prediction errors, while the ANN, despite high R^2 on the hold-out dataset, showed relatively higher RMSE and MAE values under cross-validation, reflecting sensitivity to the small dataset size.

From a transportation safety perspective, the increasing trend in pedestrian accidents with the rise in registered vehicles underscores the need for effective safety policies. These may include improved pedestrian infrastructure, speed enforcement, and public awareness targeting high-risk areas with growing motorization.

Table 6. Cross-validation performance of regression models and ANN

Model	Cross-validation	RMSE	MAE
Linear	3-fold	15,132	11,804
Linear	5-fold	14,264	11,476
Linear	7-fold	14,344	11,725
Logarithmic	3-fold	11,828	9,301
Logarithmic	5-fold	11,394	9,362
Logarithmic	7-fold	11,554	9,167
Power	3-fold	18,170	13,918
Power	5-fold	17,965	14,374
Power	7-fold	18,201	14,588
ANN	3-fold	28,326	23,910
ANN	5-fold	27,495	23,499
ANN	7-fold	27,694	23,386

Artificial intelligence and machine learning in analyzing pedestrian accidents improve traffic safety and urban planning development by giving higher accuracy. On the other hand, they have some limitations: they require a considerable amunt of data to predict correctly, as well as an experienced user, and the type of data used for training may affect the accuracy of the results. Furthermore, the ability of the ANN to capture nuanced variations in the dataset suggests that machine learning should be integrated into future transportation safety monitoring systems in Jordan and similar developing countries.

This research opens the door toward the integration of AI, ML, and ANNs to improve urban planning, traffic monitoring, and traffic safety and reduce the number of accidents and their casualties in the future.

CONCLUSIONS

Based on the results of this research, the following points were concluded:

1. Pedestrian accidents contribute to about 35% of human injury accidents.
2. Age groups (3-5) years and over 60 years are considered the riskiest in pedestrian accidents.
3. Low-speed limits of (40-60) km/hour indicate the speed group with the highest pedestrian accidents.
4. A strong linear relationship was found between registered vehicles and pedestrian accidents, with classical regression models achieving high R^2 values (up to 0.937), affirming vehicle growth as a key predictor.
5. Artificial Neural Networks (ANNs) significantly outperformed traditional models (R^2 = 0.981), demonstrating their effectiveness in capturing non-linear patterns in long-term accident data.
6. Future studies should incorporate additional explanatory variables, such as road environment characteristics, enforcement levels, pedestrian facilities, and weather conditions. Including these factors would enhance the robustness and transferability of predictive models, thereby providing more comprehensive support for data-driven policy and planning.

The results of this study demonstrated that ANN-based models outperform traditional regression in predicting pedestrian accidents from vehicle registration trends, achieving higher accuracy and reliability. From a policy perspective, these findings suggest a clear pathway for authorities in Jordan to implement AI/ML-driven traffic safety monitoring. A practical starting point would be integrating predictive models, such as those developed in this research, into the existing Traffic Management Center platforms to generate early-warning indicators of rising accident risks. Authorities can begin with pilot deployments in high-risk urban corridors identified by historical accident patterns and gradually expand coverage as technical capacity grows. By leveraging routinely collected datasets (vehicle registrations, accident statistics, and GPS feeds), the implementation cost can be minimized while ensuring scalability. Furthermore, embedding these AI/ML models within Jordan's National Road Safety Strategy would provide institutional support, enabling proactive interventions, such as targeted enforcement, improved pedestrian infrastructure, and evidence-based public awareness campaigns.

REFERENCES

Al-Masaeid, H. (2009). Traffic accidents in Jordan. *Jordan Journal of Civil Engineering, 3*(4), 331-343.

Al-Omari, B.H., & Obaidat, E.S. (2013). Analysis of pedestrian accidents in Irbid City, Jordan. *The Open Transportation Journal, 7*(1).

Bashar, A.L., Ghuzlan, K., & Hasan, H. (2013). Traffic accident trends and characteristics in Jordan. *International Journal of Civil and Environmental Engineering, 13,* 9–16.

Dongare, A.D., Kharde, R.R., & Kachare, A.D. (2012). Introduction to artificial neural network. *International Journal of Engineering and Innovative Technology, 2*(1), 189-194.

Fontaine, H., & Gourlet, Y. (1997). Fatal pedestrian accidents in France: A typological analysis. *Accident Analysis & Prevention, 29*(3), 303-312.

Gálvez-Pérez, D., Guirao, B., & Ortuño, A. (2023). Analysis of the elderly pedestrian injury severity in urban traffic accidents in Spain using machine learning techniques. *Transportation Research Procedia, 71*, 6-13.

Hua, T.K. (2022). A short review on machine learning. *Authorea Preprints*.

Institute, Jordan Traffic. (2013-2022). *Traffic accidents in Jordan*. Ministry of Interior, Public Security Directorate.

Knowles, J., Smith, L., Cuerden, R., & Delmonte, E. (2012). *Analysis of police collision files for pedestrian fatalities in London, 2006-10* (No. TfL2520).

Kouabenan, D.R., & Guyot, J.M. (2004). Study of the causes of pedestrian accidents by severity. *Journal of Psychology in Africa, 14*(2), 119-126.

Koushki, P., Al-Saleh, O., Yaseen, S., & Ali, M. (2001, January). On fatal and injurious pedestrian accidents. In *Transportation Research Board 80th Annual Meeting*.

Obaidat, M.T., & Ramadan, T.M. (2012). Traffic accidents at hazardous locations of urban roads. *Jordan Journal of Civil Engineering, 6*(4), 436-447.

Ozen, M., Sayin, C.G., & Yuruk, Y. (2017). Analysis of the pedestrian accidents in Turkey. *International Journal of Engineering and Geosciences, 2*(3), 100-109.

Sethi, D., & Zwi, A. (1999). Traffic accidents: Another disaster? *The European Journal of Public Health, 9*(1), 65-67.

Shrinivas, V., Bastien, C., Davies, H., Daneshkhah, A., Hardwicke, J., Neal-Sturgess, C.E., & Lamaj, A. (2024). Integrating machine learning in pedestrian forensics: A comprehensive tool for analysing pedestrian collisions. *SAE Technical Papers*.

Silva, A.A.D., Ströher, G.R., Teixeira, H. M., Cordeiro, M.V.G., Olandoski, M., & Von-Bahten, L.C. (2022). Impact of the COVID-19 pandemic on the epidemiology of traffic accidents: A cross-sectional study. *Revista do Colégio Brasileiro de Cirurgiões, 49*, e20223364.

Ul Arifeen, S., Ali, M., & Macioszek, E. (2023). Analysis of vehicle pedestrian crash severity using advanced machine learning techniques. *Archives of Transport, 68*.

World Health Organization. (2019). *Global status report on road safety 2018*.

Yoon, J. (2025). Prediction of high-risk areas using interpretable machine learning: Based on each determinant for the severity of pedestrian crashes. *Journal of Transport Geography, 126*, 104216.