INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue VI, June 2025

Utilization of Artificial Neural Networks (Ann) in Predicting Accidents Within Maharlika Highway San Pablo City, Laguna

Patrick Louie Jay R. Federizo[*1], Marriel Bondad-Baet[2], Arhgy L. Batarlo[3], Paul Andrei Enriquez[3], Jimuel Edmon V. Landicho[3], Juliana Marie B. Pareja[3]

[1]Asst. Professorial Lecturer, School of Environment and Design – Architecture Program, De La Salle – College of Saint Benilde, Inc., Taft Avenue, Manila

[2]Faculty, Laguna College, San Pablo City, Laguna

[3]Student, Civil Engineering Department, Laguna College, San Pablo City, Laguna

DOI: https://doi.org/10.51583/IJLTEMAS.2025.140600049

Received: 18 June 2025; Accepted: 23 June 2025; Published: 09 July 2025

Abstract – This study presents an accident prediction model for three different segments along Maharlika Highway, San Pablo City, Laguna, which utilizes an Artificial Neural Network (ANN) to analyze accident-related data from the area. The dataset included parameters such as vehicle type, driver's age, license type, gender, month of occurrence, DUI of alcohol, time and day of accidents, traffic volume, and road characteristics that were used for feature selection. Before the study creation, data preprocessing was conducted to clean and organize the data. These data were trained to predict the likelihood of RTAs based on these features. The results showed that the most relevant features based on their contribution to accidents were vehicle type with 20%, age of driver with 16%, and license with 14%. The study also predicted accidents for 2024-2028 using the ANN-LSTM Prediction. The data were divided for testing, training, and validation. Segment 1 and Segment 3 showed the lowest MSE values of 1.92 and 1.95, respectively, with an epoch of 1100, while Segment 2 showed the lowest MSE of 1.81 with 600 epochs. The LSTM prediction demonstrated different accident trends for the three segments. For Segment 1 and Segment 2, there was an increase in accidents in 2025 before peaking in 2026 and 2027, respectively, followed by a decline in the following years. For Segment 3, there was a potential drop in 2025 before peaking in 2026 with 28 predicted accidents. The predicted trends implied a possible rise in accidents in the next few years, followed by a decline.

Keywords – Accident Prediction, Artificial Neural Networks (ANN), Long-Short Term Memory (LSTM).

I. Introduction

Transportation plays a critical role in economic and social development, facilitating the movement of goods and people. However, despite its benefits, transportation is also associated with significant risks, particularly in the form of road traffic accidents (RTAs). Road traffic accidents account for approximately 1.19 million deaths annually, with an additional 20 to 50 million people suffering non-fatal injuries that often result in long-term disabilities. These accidents disproportionately affect children and young adults aged 5 to 29, with two-thirds of fatalities occurring among working-age individuals between 18 and 59 years old. The economic impact is also substantial, costing countries an estimated 3% of their gross domestic product (GDP) due to healthcare expenses, productivity losses, and other socioeconomic consequences [1].

The Philippines is not exempt from this global challenge. According to the Philippine statistics authority (PSA), road traffic accidents claimed approximately 12,000 lives in 2022 alone. Factors contributing to these incidents include poor road conditions, driver behavior, vehicle maintenance issues, and inadequate infrastructure. Additionally, developing regions such as Calabarzon exhibit higher fatality rates due to road accidents [2]. Also, it was reported an increasing number of motorcycle-related incidents, emphasizing the urgency of road safety interventions [3].

In the context of San Pablo City, Laguna, a major urban center within Calabarzon, the Maharlika highway serves as a critical transport corridor linking laguna to Quezon province. This highway has been identified as an accident-prone area, with statistics from the San Pablo City police station indicating a total of 395 road traffic accidents from 2020 to 2024. In 2024 alone, San Pablo City, along with Santa rosa city, pila, and Calamba city, recorded a high number of road accidents, indicating a continuing trend.

Traditional statistical methods used to predict road accidents often fall short in accurately capturing the complex interactions between multiple influencing factors such as weather, traffic volume, road conditions, vehicle types, and driver behavior [4]. To address this gap, artificial neural networks (ANN), a machine learning algorithm inspired by the human brain have been used effectively in various domains, including transportation engineering. ANN models excel in identifying complex patterns and relationships within large datasets, enabling highly accurate predictions [5].

Past studies have demonstrated the usefulness of ANN in solving complex problems in civil and transportation engineering [6][7]. Some research compared the use of python and r in developing ANN models for accident severity prediction, concluding that python offered higher prediction accuracy, better data visualization, and faster execution time [8].

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue VI, June 2025

Given the rising incidence of road traffic accidents and the limitations of traditional prediction methods, this study aims to develop an ANN-based system for predicting accident occurrences along Maharlika highway in San Pablo City, Laguna. The study aims to identify key parameters that contribute to road traffic accidents, train and validate an accurate ANN model, and forecast accident trends for the next five years. This prediction system is expected to aid local authorities and government agencies in strategic planning, policy development, and implementation of safety measures.

The output of this study will be particularly significant to various government agencies such as the department of public works and highways (DPWH), Philippine national police (PNP), and local government units can use the findings to improve infrastructure, enhance enforcement, and allocate resources more efficiently. Traffic authorities may use the prediction data to identify high-risk zones and implement safety interventions, such as installing additional signages and traffic calming measures.

Barangay officials may leverage localized insights to formulate community-level initiatives that target the root causes of road traffic accidents in their jurisdictions. Motorists stand to benefit from safer roads and increased awareness of accident-prone areas, while commuters may be able to make safer route decisions based on predicted risks.

Finally, the results of this research may serve as a valuable reference for future studies in the field of transportation engineering and machine learning applications, fostering further innovation in road safety interventions.

II. Methodology

In order to visualize the process behind the study, a framework was created as depicted in Figure 1.1. The study concentrated on the Maharlika Highway in San Pablo City, Laguna, using road accident reports obtained from the Philippine National Police (PNP). These reports included details such as the type of vehicle involved, roadway characteristics (like road width, length, number of lanes, and surface type), traffic volume (measured by AADT and VKT), date and time of the accident, and human-related factors (such as age, gender, license status, and alcohol influence). The number of accidents from 2020 to 2024 was also included in the analysis.

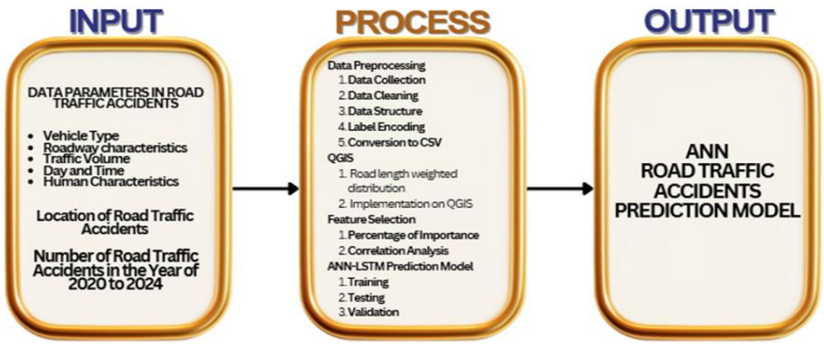


Figure 1.1 Conceptual Framework

The process began with data preprocessing to prepare for coding and analysis. This included collecting, cleaning, and structuring the data, converting categorical information into numerical values using label encoding, and saving the dataset into a CSV format.

Next, spatial analysis was performed using QGIS software to locate accident hotspots. The researchers used road length-based weighting to plot accident occurrences on a map and generate a heatmap of high-risk areas.

After this, feature selection was done to determine which data parameters had the most significant impact on accident prediction. Correlation analysis and importance ranking helped refine the input data for the model.

Finally, an Artificial Neural Network with Long Short-Term Memory (ANN-LSTM) was developed and trained using the processed data. It underwent training, testing, and validation to avoid overfitting and ensure its accuracy in predicting future road traffic accidents.

The final output was a functional predicting accident using ANN that was capable of forecasting potential accident occurrences based on historical data and identified parameters.

Equations for Feature Selection

Cramer's V Correlation

The formula of Cramer's V, implemented in the code in feature selection, was used to determine the association of the categorical parameters with the number of RTAs [9].

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue VI, June 2025

$$V = \sqrt{\frac{x^2}{N(k-1)}} \quad \text{Equation (1.1)}$$

Bayes Theorem

The researchers used Bayes' Theorem as a statistical tool to compare the results from feature selection and observational data. Bayes' Theorem determines the probability of an event based on past occurrences.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad \text{Equation (1.2)}$$

Equations for Block Spots Area

Road Length Weighted Method

A distribution is a technique that assigns weights based on their length [10].

$$Ad = \frac{Lt}{At} \quad \text{Equation (1.3)}$$

Equations for Predicting Accidents

LSTM Formula

This formula was implemented and trained in the code. It is because it was a type of recurrent neural network trained with backpropagation through time that will help in forecasting [11].

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f)$$
$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i)$$
$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o)$$
$$c_t = f_t o c_{t-1} + i_t o \sigma_c(W_c x_t + U_c h_{t-1} + b_c)$$
$$h_t = o_t o \sigma_h(c_t)$$

Equation (1.4)

Training, Testing & Validation Metrics

These formulas were implemented in the system to provide the accuracy of predicting accidents using ANN [12].

$$R^2 = \mathbf{1} - \frac{\sum(Yi-\widehat{Yt})^2}{\sum(Yi-\bar{Y})^2} \quad \text{Equation (1.5)}$$

$$MAPE = \left|\frac{x-y}{x}\right| x100 \quad \text{Equation (1.6)}$$

$$MAE = \frac{1}{n}\sum|x-y| \quad \text{Equation (1.7)}$$

$$MSE = \frac{1}{n}\sum(x-y)^2 \quad \text{Equation (1.8)}$$

$$RMSE = \sqrt{\frac{1}{n}\sum(x-y)^2} \quad \text{Equation (1.9)}$$

Research Design

This study was conducted using a quantitative research design whereby codes that implement statistical methods were utilized to study historical data sets and generate numerical data, which were quantitatively oriented. The identified parameters of the study included (a) vehicle type; (b) roadway characteristics, such as road width, length, and number of lanes; (c) measurements of traffic volume, including annual average daily traffic (AADT) and vehicle kilometers traveled (VKT); (d) day and time factors, including when the accident occurred; and (e) human characteristics, such as age, gender, license to drive, and driving under the influence of alcohol. A quantitative research approach helped the researchers establish a correlation between the given parameters and the accident data from the past five years.

Research Locale

The study took place along Maharlika Highway in San Pablo City, which stretched 4.1 km. The road started at the SM City, San Pablo, Laguna, and ended just before the Max's Restaurant in San Pablo City, with a speed limit of 60 kph across its entire length.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue VI, June 2025

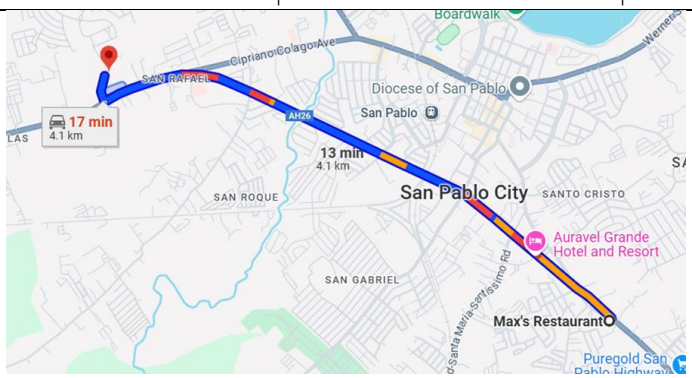


Figure 1.2 Map of Maharlika Highway, San Pablo City, Laguna (Google Maps, 2024)

Data Gathering Process

To support the development of the accident prediction framework, the researchers collected and verified various parameters to ensure accuracy before integrating them into the Artificial Neural Network (ANN). Historical data from 2020 to 2024 were gathered from credible sources like the DPWH and the PNP of San Pablo City. This included information such as accident date and time, vehicle type, and road characteristics like road width, surface type, and number of lanes. The Annual Average Daily Traffic (AADT) was converted into vehicle volume and used to calculate the Vehicle Kilometers Traveled (VKT) using standard formulas. Since the AADT for 2024 was not yet available, the researchers applied the Holt-Winters Exponential Smoothing technique to estimate values based on trends and seasonal patterns. Road capacity data were taken from DPWH standards and used to assess traffic flow and level of service on various road segments.

In addition to historical data, the researchers gathered real-time observational data through video recordings and CCTV footage at selected locations in San Pablo City. They measured AADT and classified vehicle types during peak hours over one week. This allowed the researchers to compute traffic volumes and VKT using equivalent passenger car units (PCU) for each vehicle type. Road measurements such as width and number of lanes were validated with data from DPWH. To analyze accident location, the researchers used the Road Length Weighted method, which distributed accident occurrences per barangay proportionally based on total road length, compensating for the lack of information on the location details in the police reports.

Research Instrument

The researchers used Google Colab. It allowed the installation of Python libraries in their latest versions, as it was a cloud-based platform that enabled researchers to code directly and share their work with others. It was free with a user-friendly interface and had new features, such as error detection, error explanation, and recommendations for solutions to errors. The requirements included a browser (preferably Google Chrome), internet speed of 10 Mbps or higher, a RAM 8 GB of RAM or more, and storage of 10GB for temporary files [13].

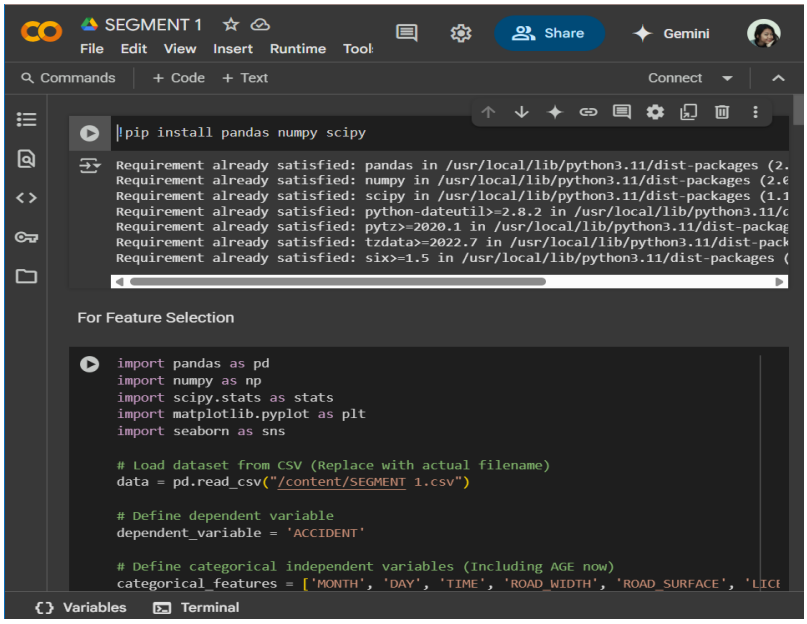


Figure 1.3 Google Colab User Interface

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue VI, June 2025

III. Results and Discussion

Accident per segment

SEGMENT	TOTAL NUMBER OF VEHICULAR ACCIDENT					
	2020	2021	2022	2023	2024	TOTAL
1	54	48	40	12	10	164
2	21	28	25	14	13	101
3	26	36	35	10	22	129
TOTAL	101	112	100	36	45	394

Table 1.1 Total number of road traffic accident in terms of year and segments

The table 1.1 shows the total number of vehicular accidents from 2020 to 2024 across three segments. Segment 1 had the highest number of accidents, with 54 in 2020, however it decreased to 10 in 2024, Segment 2 experienced 28 accidents in 2021 then decreased to 13 in 2024. Similarly, Segment 3 recorded 26 accidents in 2020, rising to 36 in 2021, before decreasing to 11 in 2023, then a slight increase to 22 in 2024.

Block spot area

Block spot area shows the segments that are prone to road traffic accidents.

Table 1.2 Number of Road Traffic Accident Per Barangay in 2020 to 2024

ROAD TRAFFIC ACCIDENT IN 2020 to 2024		
SEGMENT 1	BRGY. SAN RAFAEL	127
	BRGY. STA MONICA	2
	BRGY. SAN ROQUE	36
SEGMENT 2	BRGY. 2A	42
	BRGY. 2B	13
	BRGY. 2C	19
	BRGY. 2E	18
	BRGY. 2F	8
SEGMENT 3	BRGY. SAN FRANCISCO CALIHAN	126
	BRGY. SAN GREGORIO	3

Table 1.2 gives the data on road traffic accidents within the area of the barangays along with the three segments. Segment 1 has the highest number of accidents in Brgy. San Rafael with 127, followed by Brgy. San Roque with 36, while Brgy. Sta Monica recorded only 2 accidents. In Segment 2, Brgy. 2A with 42 accidents that makes the highest number in the segment, while Brgy. 2F with 9 accidents that were the lowest. Meanwhile in Segment 3 shows Brgy. San Francisco Calihan with 126 accidents, making it the highest in segment 3, while Brgy. San Gregorio only has 3 accidents.

Road parameters

Road Characteristics

Table 1.3 Road Characteristics Parameters

SEGMENT	ROAD LENGTH	ROAD WIDTH	NO. OF LANES	ROAD SURFACE
1	1200m	20	4	ASPHALT
2	1100m	20	4	ASPHALT
3	700m	20	4	ASPHALT

Table 1.3 illustrates the road characteristics of the three segments where the study was conducted.

Traffic Volume

Table 1.4 Level of Service relative to traffic volume

Segment	Year	Volume	Capacity	Volume/Capacity	LOS
1	2020	1,779.90	4800	0.37	B
	2021	2,329.50	4800	0.49	B
	2022	2,428.00	4800	0.51	C
	2023	2,414.60	4800	0.50	B
	2024	2,576.30	4800	0.54	C
2	2020	3,384.50	4800	0.71	D
	2021	2,764.20	4800	0.58	C
	2022	3,050.20	4800	0.64	C
	2023	2,817.20	4800	0.59	C
	2024	3,075.50	4800	0.64	C
3	2020	3,209.20	4800	0.67	C
	2021	3,198.80	4800	0.67	C
	2022	3,672.30	4800	0.77	D
	2023	3,219.80	4800	0.67	C
	2024	3,465.90	4800	0.72	D

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue VI, June 2025

Table 1.4 shows the value of Level of service per segment in the 2020-2024, in traffic engineering level of service representing how well a road operates from a traveler's perspective. In assessing traffic flow, it shows a level of service from A-F as A representing traffic of free flow and F as the worst case of traffic.

Table 1.5 Interpretation of Level of Service

Level of Service	VCR	Description
A	<0.20	Free-flow traffic
B	0.21-0.50	Relatively free-flowing traffic
C	0.51-0.70	Moderate traffic
D	0.71-0.85	Approaching unstable flow
E	0.86-1	Unstable flow
F	>1.0	Force or breakdown flow/heavy congestion

Observational Data

The observational data was data that was implemented for the result of the feature selection to know the probability of Road Traffic Accident in detailed categorical through its parameter.

Traffic volume of 2025

Table 1.6 AADT of 2025

AADT of 2025				
VEHICLE TYPE	SEGMENT 1	SEGMENT 2	SEGMENT 3	TOTAL
Motor-Tricycle	5086	7867	10664	23617
Motorcycle	6583	10089	13577	30248
Passenger Car	4087	4880	5665	14633
Passenger Utility	179	798	1417	2394
Goods Utility	1235	1542	1850	4626
Small Bus	1140	1230	1310	3680
Large Bus	558	558	558	1675
Truck (2 axle)	1311	1272	1213	3796
Truck (3 axle)	914	1016	1098	3028
Truck Semi Trailer (3 and 4 axles)	933	841	748	2522
Truck Semi Trailer (5 axles)	216	215	214	645
Truck Trailers (4 axles)	141	141	141	422
Truck trailers (5 axles)	108	108	108	325
TOTAL	22492	30556	38563	91610

The table 1.6 shows the projected Average Annual Daily Traffic (AADT) for the year 2025 across three road segments, categorized by vehicle type. Segment 3 has the highest total traffic volume with 38,563 vehicles, followed by Segment 2 with 30,556, and Segment 1 with 22,492, summing to a total of 91,610 vehicles. Motorcycles and motor-tricycles represent the largest vehicle groups, with totals of 30,248 and 23,617 respectively, indicating their dominant presence on these roads. Passenger cars also contribute significantly with 14,633 vehicles. Utility vehicles, buses, and trucks which were categorized by axles account for smaller but important portions of the traffic, reflecting the diversity of vehicles on these segments.

Table 1.7 The estimated type of vehicle crossing within the Maharlika Highway

Vehicle Type	SEGMENT 1	SEGMENT 2	SEGMENT 3	Number of Vehicle
Motor-Tricycle	2034	3147	4265	22042
Motorcycle	2633	4035	5431	28231
Passenger Car	4087	4880	5665	34143
Passenger Utility	119	532	945	3725
Goods Utility	823	1028	1233	7197
Small Bus	760	820	873	5724
Large Bus	279	279	279	1954
Truck (2 axle)	656	636	606	1898
Truck (3 axle)	366	406	439	1211
Truck Semi Trailer (3 and 4 axles)	373	336	299	1009
Truck Semi Trailer (5 axles)	86	86	86	258
Truck Trailers (4 axles)	56	56	56	169
Truck trailers (5 axles)	43	43	43	130
TOTAL	12317	16285	20222	107691

Table 1.7 shows the daily average number of vehicles going in and out through the three selected segments. It was classified by the researchers per vehicle type. In collecting the traffic volume count for the year 2025 that will be used in feature selection a 7-day and 7 hours per day volume count was done by the researchers. The researchers used CCTV footage from each segment. As it highlighted that the Passenger cars (34,143.33) have the highest estimated count, indicating that they were the most dominant vehicle type and the truck trailers, with 5 axles, had a relatively lower number of 130.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue VI, June 2025

Percentage of Importance of the parameters

This step determines the most relevant variables from data sets using the codes from python. This study has 13 total parameters that were put into training to determine their percentage of importance. Road width, road length and number of lanes were constant from the data giving no significant effect in the selection.

Table 1.8 Percentage of importance across segments

OVERALL HIGHWAY	
PARAMETERS	PERCENTAGE OF IMPORTANCE
VEHICLE TYPE	66%
AGE	54%
LICENSE	46%
GENDER	41%
MONTH	33%
DUI ALCOHOL	30%
TRAFFIC VOLUME	24%
TIME	23%
DAY	21%

The Table 1.8 illustrates the relative importance of various parameters in contributing to road traffic accidents (RTAs) in all three segments, expressed as a percentage. Vehicle Type holds the highest percentage at 66%, indicating a strong association with RTAs. This suggests that the type of vehicle significantly influences accident occurrence. Age follows with 54%, also reflecting a strong association, highlighting the critical role of the driver's age in traffic incidents. The parameters License with 46% and Gender with 41% show a moderate to relatively strong association, implying a notable influence on RTAs. Months with 33% and DUI Alcohol with 30% fall under the moderate association category. Meanwhile, Time with 23%, Day with 21%, and Traffic Volume with 24% show weak to moderate associations, indicating these factors have a comparatively lesser, though still relevant, impact on accident frequency. This overall analysis underscores that vehicle and driver characteristics are the most influential contributors to RTAs across highways.

Vehicle type

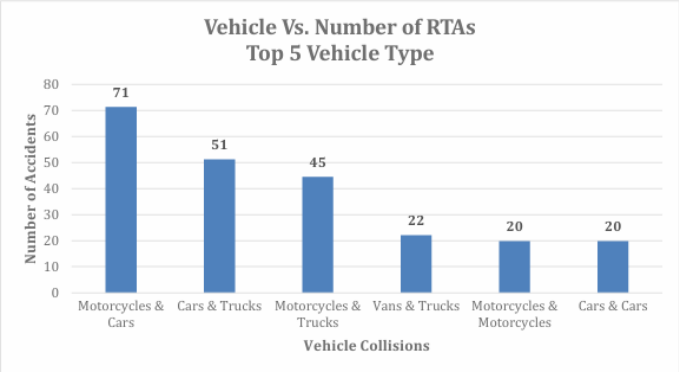


Figure 1.4 Combination of Vehicle type involved vs. Number of Road Traffic Accidents

Figure 1.4 shows what combinations of vehicle types mostly affect accidents. It presented that the combination of motorcycle and cars were the most frequently involved vehicle in an accident with 71 total accidents. Next were the combination of cars and trucks with 51 total accidents. This parameter shows that it was an important variable in the study since it contributes significantly to an accident.

Roadway Characteristics

Roadway characteristics such as road surface, number of lanes, road length, and road width shows nonimportance in the feature selection since these variables have constant values.

Traffic Volume

Table 1.9 Probability of Road Traffic Accident In terms of Traffic Volume

SEGMENT	AADT	VKT	Probability of Accident
1	22,492	26,690	42%
2	30,556	26,994	26%
3	38,563	33,612	33%

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue VI, June 2025

The table 1.9 presents data on segments on its AADT and VKT and probability of RTA for different traffic volumes. Segment 1 has the lowest Traffic Volume AADT (22,492) and VKT (26,690) but the highest accident probability at 42%. then Segment 2, with 30,556 AADT and 26,994 VKT, has the lowest accident probability at 26%, and Meanwhile, Segment 3, which has the highest AADT (38,563) and VKT (33,612), has a 33% accident probability.

Day and Time

Table 1.10 Days Vs Number of Road Traffic Accidents

Segment	Days	Number of RTAs
1	MONDAY	24
	TUESDAY	14
	WEDNESDAY	25
	THURSDAY	21
	FRIDAY	22
	SATURDAY	**40**
	SUNDAY	18
2	MONDAY	12
	TUESDAY	13
	WEDNESDAY	6
	THURSDAY	**21**
	FRIDAY	20
	SATURDAY	13
	SUNDAY	17
3	MONDAY	12
	TUESDAY	**29**
	WEDNESDAY	25
	THURSDAY	15
	FRIDAY	10
	SATURDAY	20
	SUNDAY	17

Table 1.10 shows the most frequent days for each segment where accidents occur. For segment 1, Saturday was the most frequent day with 40 accidents and for segments 2 and 3 were Thursday and Tuesday with 21 and 29, respectively.

Table 1.11 Number of accidents per segment in relation with the time of occurrence

Segment	Time	Number of RTAs
1	**Morning**	**49**
	Afternoon	41
	Evening	34
	Night	40
2	Morning	25
	Afternoon	17
	Evening	**39**
	Night	21
3	Morning	28
	Afternoon	35
	Evening	**36**
	Night	30

Table 1.11 shows the most frequent time for each segment where accidents occur. For segment 1, morning was the most frequent time with 49 accidents and for segments 2 and 3 were both evening with 39 and 36, respectively.

Human Characteristics

Table 1.12 Interpretation of Label Encoding for Human Characteristics

Driver Involved	Interpretation
Driver 1	Only one vehicle was involved in the accident, and it had one driver.
Driver 1, Driver 2	Two vehicles were involved in the accident, each with a driver.
Driver 1, Driver 2, Driver 3	Three vehicles were involved in the accident, each with a driver.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue VI, June 2025

The table 1.12 shows how to interpret the number of drivers involved in an accident. If only Driver 1 is listed, it means that only one vehicle was involved in the accident. If Driver 1 and Driver 2 are listed, then two vehicles were involved, each with its own driver. If it includes Driver 1, Driver 2, and Driver 3, then three vehicles were involved, each driven by a different person.

Table 1.13 Top 10 of Road Traffic accidents in relation with age

Top 10 Age of the driver VS. number of RTAs	
AGE	RTAs
Adult, Adult	140
Adult	40
Young Adult, Adult	40
Adult, Older Adult	21
Adult, Adult, Adult	18
Young Adult, Older Adult	17
Minor, Young Adult, Older Adult	10
Young Adult, Young Adult	8
Young Adult	7
Adult, Unidentified	7

Table 1.13 shows the most frequent combination of ages for the whole highway covering the three segments that affects the occurrence of accidents, and it shows that adult-adult was the most frequent combination of ages that affects the occurrence of accidents with 140 accidents then the second combination with the highest number was adult with 40 accidents.

Table 1.14 Age of the Driver involved in RTA for Label Encoding Interpretation

Age	Interpretation
<17	Minor
18-24	Young Adult
25-64	Adult
>65	Older Adult
No record	Unidentified

Table 1.15 Overall Number of accidents per segment in relation with license

Segment	Most frequent combination of drivers with or without driver's license involved in the accident	Number of RTAs
Overall	YES	43
	NO	13
	YES,YES	228
	YES,NO	75
	NO,NO	5
	YES,YES,YES	28
	YES,YES,NO	4

Table 1.15 shows the most frequent combination of drivers with their license for driving for the whole highway covering the three segments that affects the occurrence of accidents, and it shows that RTAs with two drivers involved and having both a license was the most frequent combination that affects the occurrence of accidents with 228 total accidents.

Table 1.16 License of the Driver involved in RTA for Label Encoding Interpretation

Categorized	Interpretation
Yes	The driver involved in the accident has a valid driver's license.
No	The driver involved in the accident does not have a driver's license.

Table 1.17 Combination of driver's gender involved vs Number of road traffic accidents

NUMBER OF ACCIDENTS				
GENDER	SEGMENT 1	SEGMENT 2	SEGMENT 3	TOTAL
Female	1	0	1	2
Male	19	10	12	41
Unidentified	0	1	2	3
Female, Male	15	4	6	11
Male, Male	**112**	**77**	**103**	**292**
Male, Unidentified	4	3	4	11
Female, Male, Male	3	0	0	3
Male, Male, Male	11	6	0	17

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue VI, June 2025

Table 1.17 shows the most frequent combination of drivers' gender per segment that affects the occurrence of accidents, and it shows that RTA's with two male drivers involved was the most frequent combination that affects the occurrence of accidents with 112 for segment 1, 77 for segment 2, and 103 for segment 3. It also shows the most frequent combination drivers' gender for the whole highway covering the three segments that affects the occurrence of accidents, and it shows that RTAs with two male drivers involved was the most frequent combination that affects the occurrence of accidents with 293 total accidents. Next was the solo male driver with 41 accidents in total.

Table 1.18 Combination of drivers DUI of alcohol vs Number of Road Traffic Accidents

	NUMBER OF ACCIDENTS			
DUI ALCOHOL	SEGMENT 1	SEGMENT 2	SEGMENT 3	TOTAL
No	36	33	49	117
Unidentified	0	0	2	2
Yes, Yes	2	3	0	5
Yes, No	5	3	2	9
Yes, Unidentified	4	0	0	4
No, No	**109**	**59**	**65**	**234**
Yes, Unidentified, Unidentified	0	0	11	11
No, No, No	8	4	0	12

Table 1.18 shows the most frequent combination of DUI of alcohol per segment that affects the occurrence of accidents, and it shows that RTAs with two drivers involved were not DUI of alcohol that affects the occurrence of accidents with 109 for segment 1, 59 for segment 2, and 65 for segment 3. The next was a solo driver that was not DUI of alcohol with 49 for segment 1, 33 for segment 2, and 36 for segment 3. It also shows the most frequent combination of DUI of alcohol for the total highway covering the three segments that affects the occurrence of accidents, and it shows that RTAs with two drivers involved were not DUI of alcohol and with 234 total accidents. Next was the solo driver that was not also DUI of alcohol with 117 accidents.

Table 1.19 License of the Driver involved in RTA for Label Encoding Interpretation

Categorized	Interpretation
Yes	the driver is under the influence of alcohol
No	the driver is not under the influence of alcohol
Unidentified	No record

f. LSTM Time Series Analysis for ANN in Predicting Accidents

Training and Testing of the Accuracy of the Predicting Accidents. The tables indicate the performance of the LSTM on three different segments with different epochs of training. The performance was measured using two important metrics: MSE and R-Squared (R^2).

Table 1.20 Performance of the Predicted Outcome through number of Epoch in Segment 1

SEGMENT 1		
EPOCHS	MSE	R^2
1100	**1.92**	**0.14**
1000	2.42	-0.08
900	2.55	-0.14
800	3.03	-0.35
700	2.85	-0.27
600	3.35	-0.49
500	3.03	-0.35

Table 1.20 Segment 1 was presented, the best performance was epoch 1100 with a MSE of 1.92 and R^2 was 0.14, demonstrating that it can account for a small but significant amount of the variance in accident frequency. While explained variance was limited, the reasonably low MSE suggests that it was accurately predicting in this segment.

Table 1.21 Performance of the Predicted Outcome through number of Epoch in Segment 2

SEGMENT 2		
EPOCHS	MSE	R^2
1100	2.6	-0.41
1000	1.86	-0.01
900	2.55	-0.38
800	2.02	-0.1
700	2.08	-0.13
600	**1.81**	**0.017**
500	2.07	-0.12

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue VI, June 2025

Table 1.21 presents the metrics for Segment 2, where the best performance was observed at epoch 600, with an MSE of 1.81 and an R^2 of 0.017. This indicates that the use of ANN in predicting accidents explains only a very small portion of the variance in accident frequency. Although the explained variance is minimal, the MSE remains within a moderate range, suggesting that the predictions are still somewhat aligned with the actual values.

Table 1.22 Performance of the Predicted Outcome through number of Epoch in Segment 3

SEGMENT 3		
EPOCHS	MSE	R^2
1100	**1.95**	**0.13**
1000	3.9	-3.87
900	2.56	-2.2
800	2.25	-1.82
700	4.33	-4.41
600	5.84	-6.3
500	4.49	-4.61

Table 1.22 tabulated the metrics for Segment 3, the best performance was epoch 1100 with the MSE was 1.95 with an R^2 of 0.13, and it replicates similar performance as in Segment 1. The prediction using ANN has consistent predictive power in this segment, and it was indicative of its potential when used on more stable or less volatile data sets.

In summary, the prediction using ANN has a promising ability in terms of reproducing accident patterns, particularly for Segments 1 and 3. Although there was potential for improvement, existing results provide a good starting point for further improvement and refinement in subsequent studies.

Validation of the Prediction

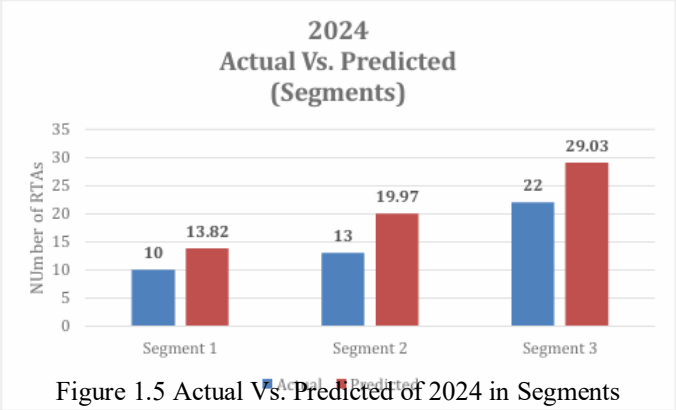


Figure 1.5 Actual Vs. Predicted of 2024 in Segments

The figure 1.5 shows the actual and the predicted road traffic accident in 2024. Given the result in segment 1 with 10 actual and 13.82 predicted. Segment 2 with 13 actual and 19.97 predicted. Lastly, segment 3 with 22 actual and 29.03 predicted.

Table 1.23 Result of MAPE, MAE, MSE, RMSE

OVERALL 2024 VALIDATION							
Segment	Actual	Predicted	Residual	MAPE	MAE	MSE	RMSE
1	10	13.82	3.82	38.2	0.851667	1.02665	1.0132
2	13	19.97	6.97	53.61538	0.964167	1.185775	1.0889
3	22	29.03	7.03	31.95455	1.245833	2.286192	1.5120
TOTAL	**45**	**62.82**	**17.82**	**41.25664**	**1.020556**	**1.499539**	**1.2245**

The Table 1.23 presents the results of the residuals, MAPE, MAE, MSE and RMSE in terms of segments. Having the predicted overlies the actual per segment having the residuals of 3.82 in segment 1, 6.97 in segment 2, 7.03 in segment 3.

In the result of the mean absolute percentage error, segment 1 with 38.2%, segment 2 with 53.61% and segment 3 with 31.95%. All in all, with a total of 41.26% which indicates a reasonable forecasting.

Then in terms of the mean absolute error, segment 1 had the lowest results having 0.85 units, following segment 2 with .96 units and segment 3 with 1.25unit. In average of 1.02 units which interprets a low value of error in the prediction using ANN.

Next was the mean squared error, with an outcome of 1.02 in segment 1, 1.18 in segment 2 and 2.29 in segment 3. On average the squared difference between the predictions and actual values was 1.5. With the result of the MSE the result of the squared root means square error had an outcome of 1.01 in segment 1, 1.09 in segment 2 and 1.51 in segment 3 with an average of 1.22 units which interprets that the accuracy had a low unit of error.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue VI, June 2025

Predicting the Number of Road Traffic Accident for the year 2024-2028 in Maharlika Highway San Pablo City

Table 1.24 Overall Predicted Number of Accidents

Overall Predicted Number of Accident				
Years	SEGMENT 1	SEGMENT 2	SEGMENT 3	TOTAL
2024	14	20	29	63
2025	37	21	18	76
2026	40	19	28	87
2027	38	19	22	79
2028	23	19	23	65

The table 1.24 depicts the total number of accidents in all of the segments within Maharlika Highway. Having the highest predicted road traffic accident was the year of 2026 with 87, following the next year of 2027 with 79 RTA, 2025 with 76, 2028 with 65 RTA and the least having a road traffic accident will be the year 2024, with 63 total of road traffic accidents.

This varying trend could be an indication of the effect of changing traffic conditions, road construction, or other external influences on accident frequencies. The declining trend towards the latter part of the forecast period could suggest the possible effect of enforced road safety initiatives, leading to a more stable and lower accident frequency.

In predicting accidents, decimal values represent the probability of an event occurring, where values between 0 and 1 indicate the likelihood of an accident, and a threshold is applied to classify the outcome.

IV. Conclusion

As one of the national roads that passes through San Pablo City, Laguna, and also due to a high number of accidents that occur, the Maharlika Highway within San Pablo City was selected as their area of study. With these, the researchers developed a long short-term memory (LSTM) and trained it to generate an accident prediction in the years of 2024 to 2028. They utilized feature-selection analysis to determine the most important variables that contribute to road traffic accidents in the selected area of study. With all the considered features, vehicle type emerged as the single most influential factor of the utilized ANN power. Human-related variables followed in importance; specifically, driver age, licensure status, and gender. Roadway characteristics such as a number of lanes, road surface, road width, and road length were constant, giving no percentage of importance in the study. Contextual or situational factors, driver alcohol impairment, month, time of day, and day of week, also held no percentage of importance in the study, validating the interaction between human action and environmental conditions. Collectively, these findings affirm that the nature of the accidents that human characteristics and external conditions influence accidents.

According to the analysis of the road traffic accident (RTA) counts along the three highway segments, identified block spot zones. In Segment 1, Barangay San Rafael registered the greatest number of RTAs and therefore constitutes the primary block spot for that segment. Segment 2 exhibited a similar concentration in Barangay 2-A, whereas Segment 3 displayed nearly equivalent accident frequencies in Barangay San Francisco Calihan. By contrast, Barangays Sta. Monica and San Gregorio recorded only sparse crash occurrences. These results highlight these segments as geographic zones that have a high accident rate and need carefully planned and targeted interventions.

The performance of the artificial neural network (ANN) for predicting road traffic accidents (RTAs) was evaluated with mean-square error (MSE) and R-squared (R^2) in different ranges of epochs ranging from 500 to 1,100 epochs. For Segments 1 and 3, the 1 100-epoch configuration yielded the lowest MSE and the largest R^2, indicating the most accurate and best-fitting predictions among all tested settings. Segment 2 achieved its optimum at 600 epochs, recording the smallest MSE for that segment. Overall, the 1,100-epoch network provided the most reliable overall predictive capability for the accident prediction study. Further validation against actual accident data from 2024 revealed minimal prediction error, as this was considered to be acceptable and reasonable for accident prediction.

The long short-term memory (LSTM) effectively predicted accident trends in the five years (2024–2028), revealing distinct patterns across the three highway segments. Segment 1 forecasts indicated an increase in accidents peaking in 2026, followed by a decline through 2027 and 2028. Similarly, Segment 2 showed a rise in 2025, stabilized in 2026, with a peak in 2027 before decreasing thereafter. In contrast, Segment 3 experienced a decline in 2025, a peak in 2026, and a gradual reduction in the last 2 years. These projections suggested a short-term increase in accident occurrences, followed by a gradual decline. Analysis of the results indicated that motor-tricycle, car, and motorcycle drivers, particularly males holding valid driver's licenses, were predominantly involved, with incidents frequently occurring during evenings and weekends. These insights provided valuable guidance for traffic enforcement agencies and policymakers aiming to design targeted interventions and public awareness campaigns to effectively reduce road traffic accidents in the study area.

V. Recommendations and Future Work

To enhance the predictive accuracy, interpretability, and practical relevance of the ANN-LSTM model used in this study, the following improvements are recommended for future research:

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XIV, Issue VI, June 2025

Improve Model Interpretability: To better understand how individual features contribute to the prediction of road traffic accidents (RTAs), it is recommended to incorporate feature importance visualizations such as SHAP (SHapley Additive exPlanations) values. These tools provide a more intuitive explanation of how each variable influences the model output, enabling clearer interpretation by policymakers and researchers. Additionally, comparative analysis with conventional statistical models such as Poisson regression and logistic regression is advised. This benchmarking will help validate the robustness and practical advantages of the ANN-LSTM model.

Expand Data Collection and Validation: Future iterations of this study should integrate external and environmental variables including real-time weather conditions (e.g., rainfall, visibility), road lighting conditions, and pedestrian activity. These contextual factors are known to significantly influence accident rates. Moreover, validation should extend beyond historical datasets by incorporating real-time accident data for years beyond 2024, ideally through automated feeds from law enforcement or transportation authorities.

Clarify Methodology and Visualization of Results: To enhance replicability, future papers should include pseudo-code or a detailed flowchart that outlines the steps in the ANN-LSTM model development. Additionally, the inclusion of comprehensive visualizations—such as heatmaps for accident density, time-series plots for trend analysis, and confusion matrices for classification performance—can substantially improve the interpretability of results.

Discuss Limitations More Explicitly: It is important to acknowledge potential biases inherent in police-reported accident data, particularly the underreporting of minor incidents. Furthermore, the model should account for seasonal variations in traffic patterns, including increased activity during holidays or weather-related anomalies such as monsoons, which may impact driving behavior and accident frequency.

Policy and Safety Recommendations: Based on the predicted accident trends, this study supports the implementation of targeted safety interventions, such as: (1) Installation of speed cameras and automated enforcement in high-risk zones. (2) Enhanced DUI enforcement during peak evening and weekend hours. (3) Improved road lighting and reflective signage in accident-prone barangays.

A cost-benefit analysis of these proposed interventions should be conducted to ensure resource-efficient implementation. The estimated reduction in accident frequency and severity may justify the investment, particularly when weighed against public health costs and productivity losses due to traffic incidents.

References

1. World Health Organization: WHO. (2023). Road traffic injuries. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries
2. Velasco, V., Robielos, R., & Gumasing, Ma. J. (2021). Analysis of road traffic accident criticality in CALABARZON, Philippines.
3. Golden Future Life Plans. (2024). Top 10 causes of death in the Philippines 2023. https://www.goldenfuture.com.ph/articles/top-10-causes-of-death-in-the-philippines-2023/.
4. Mannering, F., Bhat, C. R., Shankar, V., & Abdel-Aty, M. (2020). Big data, traditional data and the tradeoffs between prediction and causality in highway-safety analysis. Analytic Methods in Accident Research, 25, 100113.https://www.sciencedirect.com/science/article/abs/pii/S2213665720300038?via%3Dihub.
5. Qamar R. & Zardari B.A. (2023). Artificial Neural Networks: An Overview. Mesopotamian journal of Computer Science. Vol. (2023), 2023, pp. 124–133. https://mesopotamian.press/journals/index.php/cs/article/view/118/121.
6. Marzouk, M., Elhakeem, A., & Adel, K. (2023). Artificial neural networks applications in construction and building engineering (1991–2021): Science mapping and visualization. Applied Soft Computing, 152, 111174. https://doi.org/10.1016/j.asoc.2023.111174
7. Soori, M., Arezoo, B., & Dastres, R. (2023). Artificial neural networks in supply chain management, a review. Journal of Economy and Technology, 1, 179–196. https://doi.org/10.1016/j.ject.2023.11.002
8. Dipto, I. C., Rahman, M. A., Islam, T., & Rahman, H. M. M. (2020). Prediction of accident severity using artificial neural network: A comparison of analytical capabilities between Python and R. *Journal of Data Analysis and Information Processing, 08(03) *, 134–157. https://doi.org/10.4236/jdaip.2020.83008
9. Zhang, Y., Ren, G., Liu, X., Gao, G., & Zhu, M. (2022). Ensemble learning-based modeling and short-term forecasting algorithm for time series with small sample. Engineering Reports, 4(12), e12486. https://doi.org/10.1002/eng2.12486.
10. Ganaie, R. A., & Rajagopalan, V. (2021). Length biased weighted new quasi Lindley distribution: Statistical properties and applications. Pakistan Journal of Statistics and Operational Research, 17(1), 123-136. https://pjsor.com/pjsor/article/view/3084/1106.
11. Python Operating Systems List. (2025). Python.org. https://www.python.org/downloads/operating-systems/.
12. Hyndman, R. J., & Athanasopoulos, G. (2021). Forecasting: Principles and Practice (3rd ed.). OTexts. https://otexts.com/fpp3/
13. Das, T. (2025). Google Colab: What, how, and features. Geekflare. https://geekflare.com/dev/google-colab/