

Integrating Statistical, Machine Learning, and Deep Learning Approaches for Urban Traffic Accident Forecasting

Prof. Sangita Kansal

Professor (Graph Theory and Petri Net Theory),

Department of Applied Mathematics

Delhi Technological University (formerly Delhi College of Engineering), Delhi, India

sangita_kansal15@dce.ac.in

Aamir Sayeed

Student,

Department of Applied Mathematics

Delhi Technological University (formerly Delhi College of Engineering), Delhi, India

aamirsayeed_mc22a13_26@dtu.ac.in

Arun Bari

Student,

Department of Applied Mathematics

Delhi Technological University (formerly Delhi College of Engineering), Delhi, India

arunbari_mc22a13_52@dtu.ac.in

Rohan Kamewar

Student,

Department of Applied Mathematics

Delhi Technological University (formerly Delhi College of Engineering), Delhi, India

rohankamewar_mc22a14_75@dtu.ac.in

Abstract— During rush hours, urban mobility is significantly impacted by traffic accidents, incurring significant economic and human costs, and requiring accurate forecasting tools to make proactive decisions. The complex spatial and temporal structures observed in real-world accident data are often complex and challenging to model using traditional statistical techniques. To address these problems, an Accident Prediction dataset from Kaggle is used as a publicly accessible dataset and a time-series forecasting framework is created for predicting the peaks of traffic accidents. The proposed preprocessing steps are dependent on location and peak hour characteristics, and involve cleaning, normalization, categorical encoding, exploratory analysis, and feature engineering. A number of prediction models are used, such as a long short-term memory network to capture long-range temporal patterns, a multilayer perceptron with lagged inputs, SARIMAX with exogenous variables, and a Negative Binomial generalized linear model. The model's performance is evaluated using mean error, mean absolute error, root mean square error, mean absolute percentage error and mean absolute scaled error. Based on the results, the LSTM model has the lowest MAE, approximately 1.35, and RMSE, approximately 1.85, making it the most accurate in prediction. LIME and SHAP are added to enhance the explainability, while the addition of a user-driven forecasting interface using Flask contributes to increased accuracy and usefulness of the prediction in the field of urban traffic safety.

“Keywords— Multi-Layer Perceptron (MLP), Negative Binomial (NB), Pearson Correlation Analysis, Prediction Model, Seasonal Auto Regressive Integrated Moving Average (SARIMA-X), Traffic Accident Rate”.

I. INTRODUCTION

Road traffic accidents are a major public health and socio-economic problem that causes a high number of deaths, injuries and property damage around the world. Rapid urbanization, increased motorization, and complex traffic interactions, particularly during rush hours when the demand for traffic is highest, have increased the risk of accidents. [1] Road safety plays a key role in sustainable development and international health and transportation

organisations have consistently called for data-driven solutions to prevent the incidence and impact of accidents [2]. Consequently, many countries have adopted long-term plans and regulations for road safety, aiming to reduce the damage that traffic causes by systematic research and preventive measures [3].

Despite these efforts, traffic accidents are difficult to predict reliably, because of the randomness of what happens and the varying interactions between traffic flow, human behavior, and highway conditions. Some studies have studied accident prediction through different modelling approaches, often focusing on specific space or time units [4, 5]. On the interpretation of temporal accident patterns by various modeling viewpoints, however, there is little agreement, especially at fine-grained time resolutions pertinent to peak hours [6]. Additionally, the lack of comparative knowledge about the relative merits and deficiencies of various model types for traffic safety analyses is caused by the absence of methodological knowledge in terms of how well they capture the variability of events over time.

This study's main goal is to provide a thorough analytical framework for predicting traffic accidents during peak hours utilizing accident data from the past that has been enhanced with pertinent traffic and highway characteristics. This study aims to examine and compare the prediction properties of the different approaches to the sample problem of various modeling paradigms with the context of treating accident occurrences as time-series processes. The study seeks to offer a comprehensive view of the temporal accident forecasting through a systematic analysis of model outputs under the same sets of data, thus gaining more understanding on the influence of temporal dynamics and contextual features on the process of accident prediction [8].

This research is important because of potential applications to proactive traffic management and evidence-based roadway safety planning. Accurate peaks hour accident projections can help transportation organisations identify high-risk periods, optimally allocate resources and

develop targeted safety interventions. Furthermore, the perspective of comparing accident investigation approaches in the study contributes to the advancement of methodology; this perspective includes educating academics and practitioners on proper modeling methods for time-series accident investigation. The ultimate goal of the results is to bring policies and other decision makers towards achieving long-term road safety goals and reducing the burden of traffic accidents on society [9, 10].

II. RELATED WORK

Early research on traffic accident modeling has concentrated on statistical methods for crash frequency and to understand underlying traffic accident risk factors. Quddus [11] discussed time-series count data models for traffic accidents, which he demonstrated to be suitable for traffic crash data. Although useful for long-term analysis, the study brought to light issues with overdispersion and restricted adaptability when depicting intricate traffic dynamics. More recently, Roland et al. [12] showed that data-driven methods can improve the predictive power by modelling the problem from a temporal perspective, predicting the number of vehicle accidents at a city level. However, the study was limited and focused, and offered only a small amount of comparison between different modeling paradigms.

Besides the aforementioned, time-series forecasting has also been researched to assess the impact of external factors on traffic safety. Sekadakis et al. [13] showed how time-series models can be used to detect abrupt structural changes in accident patterns during the COVID-19 pandemic. The national-level aggregation despite the benefits of the study may have hindered the visibility of regional and peak hour variation which is important for operational safety planning. Wang et al. [14] reported that temporal resolution plays a significant role in model performance when the safety models are compared at different traffic aggregations levels. Their investigation however failed to examine the performance of different classes of models in the same time-series context.

Risk analysis research has provided additional model-related insights beyond those provided by traditional traffic safety research. Shikder et al. [15] discussed the importance of operational and contextual factors in predicting safety measures in the context of hazardous materials transportation risk incidents. These studies are useful but often tend to focus on risk assessment rather than predicting when there will be a commotion on the streets. Geographical heterogeneity is recognized as a problem in crash modelling and Xu and Huang [16] demonstrated that incorporating location-specific variability improves the amount of explanatory power. But methods that are only spatially based may miss important temporal dependencies that are important for making predictions about accidents during peak hours.

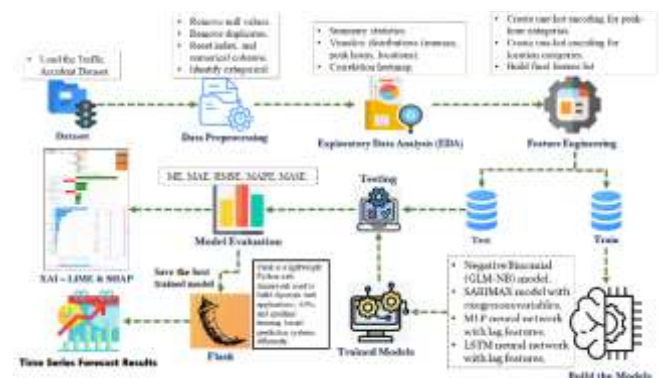
State-by-state analyses have also underscored the importance of local data to safety modeling. The need for context-aware prediction frameworks is reinforced by the work conducted by Young and Park [17], who consider and showcase the superiority of municipality-specific safety performance functions over generalized safety performance functions. In earlier work, Joshua and Garber [18] introduced regression-based methods to estimate accident rates, which

were simple but lacked the ability to adapt to nonlinear and time-varying patterns. In the same way, Jovanis and Chang [19] were able to establish accident correlations through exposure measures such as miles traveled. The model provided valuable insights into risk normalisation, even though it was based on relatively fixed assumptions.

Finally, methodological compromises have been found by comparing count regression methods. Miaou [20] compared alternative 'count' based formulations with simpler ones, and found that these alternative formulations can better accommodate overdispersion in accident data. While these research may advance the state of affairs of traffic safety modeling more broadly, they often examine individual methods or focus on either the temporal or spatial aspects of traffic safety. A major gap in the systematic comparison of different modeling views remains for the fine-grained, peak hour time-series accident forecasting. The current study addresses this gap by offering a uniform comparative framework that analyses the predictive behaviors across multiple model paradigms with a common temporal characteristics using consistent temporal data.

III. MATERIALS AND METHODS

The proposed method is based on spatial-temporal characteristics of the public data of traffic accidents to predict traffic accidents for peak hours. The unified modeling pipeline transforms accident reports into a consistent time-series representation which captures the temporal dependency and location-specific behavior. Data partitioning using intersection awareness is able to provide spatial consistency, while data aggregation by lag and categorical peak-hour representations add temporal dynamics. To handle a variety of data features, such as nonlinear relationships, seasonal changes, and over-dispersed accident numbers, several complimentary prediction models are integrated. An extra recurrent forecasting module is added to further enhance long-range temporal learning, which can provide more accurate modelling of complex patterns of accident evolution. An approach to XAI that can make models predictable and measure feature influence increases model transparency. In addition, a lightweight deployment layer ensures real-time inference and ease of use. The aim of the system is to deliver accurate, understandable and deployable estimates of accidents from the peak hour to improve data-driven urban traffic safety management.



"Fig. 1. System Architecture"

After collecting data and label encoding, feature selection by the PPS approach is proposed in the suggested system design. The data is then split into training and testing sets, and plotted. Some of the ML models built and trained are evaluated using accuracy, precision, recall, F1-score and cross validation accuracy. The best trained model that performs best is saved and deployed using Flask. The approaches of XAI such as SHAP and LIME are employed to boost transparency and offer an explanation of model predictions.

A) Dataset Collection

A publicly accessible Traffic Peak-Hour Accident dataset from an open Kaggle repository is used in the study. There are 924 records in the collection, each of which is an observation of traffic accidents at a crossroad in the metropolitan area. It features a variety of numerical and categorical attributes including intersection identifiers, peak hour categories, location types, geometric characteristics, traffic volume indicators and a target variable which is the number of accidents. The data set is complete, well-structured and suitable for modelling spatial-temporal accident patterns, making it suitable for peak hour traffic accident forecasting research.

	intsec	peakhour	location	noleg	sighead	wdADTmaj	wdADTmin	numacc
0	1	mp1	South	6	2	750	305	3
1	1	mp2	South	6	2	1136	422	2
2	1	ap1	South	6	2	1722	500	12
3	1	ap2	South	6	2	1097	282	5
4	1	ep1	South	6	2	939	533	4

“Fig. 2 Traffic Accident Dataset”

B) Pre-Processing

In this part, the core data preparation methods for assuring data quality, representing features, and facilitating the trustworthy construction of a spatial-temporal model for accurate forecasting of peak hour traffic accidents are presented.

Data Inspection and Exploratory Analysis: An initial inspection and exploratory analysis of the dataset were performed to understand the structure, quality and statistical attributes of the dataset. In this phase, data types, distribution of features, central trends and variability of attributes associated with accidents were analyzed. Visual exploration was used to determine the extent of accident frequency trends by peak hours, locations, and intersections. In order to make the right preprocessing and modeling decisions, it is crucial to detect data properties including skewness, variability, and potential modeling problems.

Data Cleaning and Quality Assurance: Before modeling, data cleaning was done to guarantee the dataset's consistency and dependability. To avoid the possibility of skewing the data analysis and affecting the accuracy of the forecasting, the data set was screened for redundant data and missing data. There were no null or duplicate entries in the dataset, so it was an entire, intact dataset. A data structure was used to organize the data; index resetting was used to do this. This step is crucial for ensuring data accuracy, preventing data leaks, and ensuring the model performs consistently and reliably.

Categorical Feature Encoding: Categorical data were converted to numerical data using encoding techniques, which included the categories of the peak hour and the spatial location zones. This translation makes it possible for statistical and machine learning algorithms that need numerical inputs to efficiently use categorical features. The system can then use that data to reveal any temporal or spatial variations in accident patterns. This also boosts the expressiveness of the models while ensuring that the qualitative traffic properties have a substantial influence on the forecasting accuracy.

Correlation and Dependency Analysis: A correlation study was conducted to examine linear correlations between traffic volumes, geometric characteristics, temporal indicators and accident frequency. This stage helps to identify possible multicollinearity among characteristics and influential factors by computing pairwise correlations. Dependency structure visualization helps to understand the dependency between collision frequency and traffic and spatial variables. Comprehending these correlations facilitates well-informed feature selection, enhances the interpretability of the model, and fortifies the dependability of later predictive modeling.

C) Training and Testing:

The data set was divided into training and test sets based on different intersection identifiers to achieve spatial independence of data partitions. This makes for a model evaluation that ensures generalization to unseen junctions and prevents information leakage between sites. The method can accurately simulate the deployment environment of new metros and is able to sample a certain percentage of the junctions for testing. This stage is crucial for strong performance assessment and objective evaluation of the spatial-temporal forecasting ability.

D) Algorithms

Negative Binomial Model: The Negative Binomial model is used for the analysis of over-dispersion, accident frequency data. It provides robust and understandable count-based forecasts, while taking into account the variance that is larger than the mean level, which leads to a reliable base level of performance for simulating a range of traffic accidents.

SARIMAX Model: The SARIMAX model is a model that takes into account the seasonal variation and temporal dependence of the accident time series, and also accounts for the external factors that influence the accidents. It is structured to represent autoregressive and moving-average components which leads to short-term forecasting stability under traffic and peak-hour changes.

Multilayer Perceptron (MLP): MLP models consists of a sequence of nonlinear transformations that are able to represent nonlinear dependence between past inputs and future accident counts. When accident dynamics are nonlinear, this ability can be used to learn patterns flexibly and generalize the patterns.

Long Short-Term Memory (LSTM): The LSTM enables the network to retain relevant past information over a long time period, and to mimic sequential accident patterns. It is

resilient, and its gated memory structure enhances accuracy and prediction of dynamic accident trends, where long-range temporal dependencies exist.

IV. EXPERIMENTAL RESULTS

MAE: The degree of mistake in your measurements is known as absolute error. It's the difference between the actual value and the measured value. For example, if the scale reads 90 pounds but you know that you weigh 89 pounds, then the absolute inaccuracy is $90 - 89 = 1$ pound.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (1)$$

RMSE: RMSE is used to measure the average difference between the estimated values and the observed values of a statistical model. From the mathematical point of view, it is the standard deviation of the residuals. Residuals are the distances between the regression line and the data points.

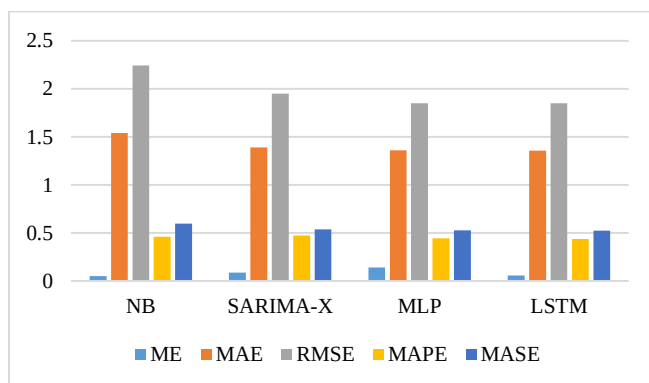
$$RMSE = \sqrt{\frac{\sum_{i=1}^n |y(i) - \hat{y}(i)|^2}{N}} \quad (2)$$

Table. 1. Performance Evaluation

Model	ME	MAE	RMSE	MAPE	MASE
NB	0.050722	1.540643	2.243663	0.460515	0.595413
SARIMA-X	0.085867	1.389943	1.947731	0.473590	0.537172
MLP	0.140733	1.357603	1.849476	0.441441	0.524674
LSTM	0.057551	1.354583	1.848039	0.434560	0.523507

As shown in the table, the table shows that the LSTM model has the lowest total prediction errors, thereby demonstrating higher forecasting accuracy and robustness in predicting peak hour accidents.

Fig. 3. Comparison Graph



In the bar chart that compares the overall error values of the forecasting models, the overall error of the LSTM model is lower than that of the other models, indicating that the error values of the forecasting models gradually decreased from the statistical models to the DL models.



Fig. 4 Enter the input data

If the user credentials entered and checked by the system match, the input result will display the user credentials that have been accepted and passed to the system, which means that the submission was successful and it is now possible to proceed with the study of traffic accident forecasts during peak hours with secure authentication and access.



Fig. 5 Predicted Result

The expected output includes predicted results based on different ML models and peak hour traffic accident risk assessments based on various inputs to support data-driven and accurate traffic safety decisions.



Fig. 6 Enter the input data

The inputs include factors like peak hour, location, traffic volume, road infrastructure and forecast horizon factors, which are then analyzed to create a web-based dashboard that predicts the likelihood of traffic accidents during peak hours, as shown in FIG. 6.



Fig. 7 Predicted Result

The PEAK HOUR output MP1 (07–08) graph displays the time-series traffic accident projections generated by Naive Bayes, SARIMAX, MLP, and LSTM models, respectively, over each of the subsequent hourly prediction intervals.

V. CONCLUSION

The idea is to address an immediate need for accurate forecasts of traffic accidents during peak hours in urban areas, where timely forecasts can assist in making safety-based decisions and planning. An extensive time-series forecasting pipeline is designed to generate spatial-temporal accident patterns using a publicly available Kaggle Accident Prediction data set. In addition to the above, the process includes data cleaning, normalization, category encoding, exploratory data analysis and feature engineering of geographical and peak hour properties. Assessment of forecasting performance is provided by a variety of modeling techniques, including a multilayer perceptron neural network that models nonlinear relationships by using lagged inputs, a SARIMAX model with exogenous variables for accounting for seasonal and temporal trends, and a Negative Binomial generalized linear model for over-dispersed count data. For assessment of performance, standard error based measures such as Mean error, Mean absolute error, RMSE, Mean absolute percentage error and Mean absolute scaled error are used. A LSTM network is further introduced to enhance the temporal modeling capability to enable effective learning of long-range temporal relationships. The improvement confirms the effectiveness of deep temporal modeling for accident forecasting, and shows that it outperforms the performance with a mean absolute error of approximately 1.35. LIME and SHAP provide an explanation of the impact of the features, making the model more interpretable, and the flask-based interface enables users to make predictions. Considering all of this, the system makes a reliable, transparent and practically feasible contribution to the field of automatic support in traffic accidents during the peak hour.

Future improvements could focus on developing forecasting resilience, incorporating additional context elements such as weather conditions, traffic flow, road

structure, and event-related disturbances, to better capture real-world unpredictability. Real-time data streams from intelligent transportation systems and traffic sensors can be used to integrate into dynamic and adaptive accident prediction. Advanced hybrid and ensemble learning methods can be explored to harness the strengths of the statistical and deep learning approaches. Also, increasing the range of explainability methods to offer spatiotemporal visual analytics could increase the understanding of stakeholders. Cloud-based implementation helps to enhance the scalability by providing pro-active decision assistance and the city-wide extension of traffic safety monitoring.

REFERENCES

- [1] Babanezhad, M., Khorsha, H., Mohajervatan, A., & Choori, A. (2025). Estimating the Demand for Ambulances in Traffic Accidents. *Health in Emergencies and Disasters Quarterly*, 10(4), 247-258.
- [2] Quest, W., Goodness, W., & Elly, B. (2025). Statistical Modeling of Road Traffic Accident Patterns and Prediction of Future Accident Hotspots. Preprint on ResearchGate.
- [3] Wu, H. (2025). The Relationship Between Traffic Flow Forecasting and Traffic Accident Forecasting and the Possible Combination Points. In *ITM Web of Conferences* (Vol. 70, p. 01006). EDP Sciences.
- [4] Mohammed, N., Ware, A. N., & Sodano Sheiso, D. (2025). Time Series Modelling of Road Traffic Accidents in West Arsi, Ethiopia. Ethiopia (April 22, 2025).
- [5] [5] Park, D., Choi, S. S., Lim, D., & Kang, Y. S. (2025). Graph Multi-Resolution Transformer for Road Traffic Anomaly Detection. *IEEE Access*.
- [6] Global Status Report on Road Safety: Time for Action, Geneva, World Health Organization. 2009, Accessed: Oct. 23, 2023. [Online]. Available: <https://www.afro.who.int/publications/global-status-report-road-safety-time-action>
- [7] Global Status Report on Road Safety, Geneva, World Health Organization. (2018). [Online]. Available: <https://www.who.int/publications/i/item/9789241565684>
- [8] The Global Goals: Good Health and Well-Being. (2017). [Online]. Available: <https://www.globalgoals.org/goals/3-good-health-and-well-being/>
- [9] Canada's Road Safety Strategy 2025. (2016). [Online]. Available: <https://roadsafetystrategy.ca/web/road-safety-strategy/files/public/docs/RSS-2025-Report-July-2016-with-cover.pdf>
- [10] C. C. Ihueze and U. O. Onwurah, "Road traffic accidents prediction modelling: An analysis of Anambra State, Nigeria," *Accid. Anal. Prev.*, vol. 112, pp. 21–29, Mar. 2018, doi: 10.1016/j.aap.2017.12.016.
- [11] M. A. Qudus, "Time series count data models: An empirical application to traffic accidents," *Accid. Anal. Prev.*, vol. 40, no. 5, pp. 1732–1741, 2008, doi: 10.1016/j.aap.2008.06.011.
- [12] J. Roland, P. D. Way, C. Firat, T.-N. Doan, and M. Sartipi, "Modeling and predicting vehicle accident occurrence in Chattanooga, Tennessee," *Accid. Anal. Prev.*, vol. 149, Jan. 2021, Art. no. 105860, doi: 10.1016/j.aap.2020.105860.
- [13] M. Sekadakis, C. Katrakazas, E. Michalaraki, F. Kehagia, and G. Yannis, "Analysis of the impact of COVID-19 on collisions, fatalities and injuries using time series forecasting: The case of Greece," *Accid. Anal. Prev.*, vol. 162, Nov. 2021, Art. no. 106391, doi: 10.1016/j.aap.2021.106391.
- [14] L. Wang, M. Abdel-Aty, X. Wang, and R. Yu, "Analysis and comparison of safety models using average daily, average hourly, and microscopic traffic," *Accid. Anal. Prev.*, vol. 111, pp. 271–279, Feb. 2018, doi: 10.1016/j.aap.2017.12.007.
- [15] M. F. H. Shikder, Y. Tang, E. Almedhawe, and J. C. Araújo, "Risk incident analyses in the transportation of anhydrous ammonia as an emerging clean energy resource," *Risk Anal.*, vol. 45, no. 3, pp. 653–667, 2025, doi: 10.1111/risa.17634.

- [16] P. Xu and H. Huang, "Modeling crash spatial heterogeneity: Random parameter versus geographically weighting," *Accid. Anal. Prev.*, vol. 75, pp. 16–25, Feb. 2015, doi: 10.1016/j.aap.2014.10.020.
- [17] J. Young and P. Y. Park, "Benefits of small municipalities using jurisdiction-specific safety performance functions rather than the Highway Safety Manual's calibrated or uncalibrated safety performance functions," *Can. J. Civ. Eng.*, vol. 40, no. 6, pp. 517–527, 2013, doi: 10.1139/cjce-2012-0501.
- [18] S. C. Joshua and N. J. Garber, "Estimating truck accident rate and involvements using linear and Poisson regression models," *Transp. Plan. Technol.*, vol. 15, no. 1, pp. 41–58, 1990, doi: 10.1080/03081069008717439.
- [19] P. P. Jovanis and H.-L. Chang, "Modeling relationship of accidents to miles traveled," *Transp. Res. Record J. Transp. Res. Board*, vol. 1068, pp. 42–51, Dec. 1986. [Online]. Available: <http://onlinepubs.trb.org/Onlinepubs/trr/1986/1068/1068-005.pdf>
- [20] S.-P. Miaou, "The relationship between truck accidents and geometric design of road sections: Poisson versus Negative Binomial regressions," *Accid. Anal. Prev.*, vol. 26, no. 4, pp. 471–482, Aug. 1994, doi: 10.1016/0001-4575(94)90038-8.
- [21] M. A. Abdel-Aty and A. E. Radwan, "Modeling traffic accident occurrence and involvement," *Accid. Anal. Prev.*, vol. 32, no. 5, pp. 633–642, 2000, doi: 10.1016/S0001-4575(99)00094-9.
- [22] T. Champahom, S. Jomnonkwao, A. Karoonsoontawong, and V. Ratanavaraha, "Spatial zero-inflated Negative Binomial regression models: Application for estimating frequencies of rear-end crashes on Thai highways," *J. Transp. Saf. Secur.*, vol. 14, no. 3, pp. 523–540, Mar. 2022, doi: 10.1080/19439962.2020.1812786.
- [23] Y.-A. Daraghmi, C.-W. Yi, and T.-C. Chiang, "Negative binomial additive models for short-term traffic flow forecasting in urban areas," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 2, pp. 784–793, Apr. 2014, doi: 10.1109/TITS.2013.2287512.
- [24] A. E. Retallack and B. Ostendorf, "Relationship between traffic volume and accident frequency at intersections," *Int. J. Environ. Res. Public Health*, vol. 17, no. 4, p. 1393, 2020, doi: 10.3390/ijerph17041393.
- [25] A. Alqatawna, A. M. R. Álvarez, and S. S.-C. García-Moreno, "Comparison of multivariate regression models and artificial neural networks for prediction highway traffic accidents in Spain: A case study," *Transp. Res. Procedia*, vol. 58, pp. 277–284, Jan. 2021, doi: 10.1016/j.trpro.2021.11.038.
- [26] C. Zegeer, "Cost-effective geometric improvements for safety upgrading of horizontal curves," *TRID*. 1991. [Online]. Available: <https://trid.trb.org/view/349220>
- [27] V. Shankar, F. Mannering, and W. Barfield, "Effect of roadway geometrics and environmental factors on rural freeway accident frequencies," *Accid. Anal. Prev.*, vol. 27, no. 3, pp. 371–389, 1995, doi: 10.1016/0001-4575(94)00078-Z.
- [28] H. C. Chin and M. A. Quddus, "Applying the random effect negative Binomial model to examine traffic accident occurrence at signalized intersections," *Accid. Anal. Prev.*, vol. 35, no. 2, pp. 253–259, 2003, doi: 10.1016/S0001-4575(02)00003-9.
- [29] A. S. M. M. Islam, M. Shirazi, and D. Lord, "Finite mixture Negative Binomial-Lindley for modeling heterogeneous crash data with many zero observations," *Accid. Anal. Prev.*, vol. 175, Sep. 2022, Art. no. 106765, doi: 10.1016/j.aap.2022.106765.
- [30] S. Mathew, S. S. Pulugurtha, and S. Duvvuri, "Exploring the effect of road network, demographic, and land use characteristics on teen crash frequency using geographically weighted Negative Binomial regression," *Accid. Anal. Prev.*, vol. 168, Apr. 2022, Art. no. 106615, doi: 10.1016/j.aap.2022.106615.