Mechatronics and Intelligent Transportation Systems
https://www.acadlore.com/journals/MITS

Adoption of Real-Time Data and Functional Modeling to Predict Urban Traffic Crashes

Izhar Ullah*

Mathematics Research Centre, Near East University, 99138 Nicosia, North Cyprus

* Correspondence: Izhar Ullah (20234585@std.neu.edu.tr)

Received: 06-20-2025 **Revised:** 07-08-2025 **Accepted:** 07-16-2025

Citation: I. Ullah, "Adoption of real-time data and functional modeling to predict urban traffic crashes," *Mechatron. Intell Transp. Syst.*, vol. 4, no. 3, pp. 134–142, 2025. https://doi.org/10.56578/mits040303.

Abstract: This study presented a novel mathematical functional-based algorithm designed to predict the risks of vehicular crashes by leveraging real-time traffic data collected from urban road networks. The proposed model integrated multiple critical variables, including traffic speed, vehicle density, visibility conditions, spatial coordinates, and time-of-day factors, to generate a comprehensive and dynamic assessment for foreseeing the likelihood of traffic crashes. The flexible functional framework enabled the incorporation of diverse traffic and environmental variables, thereby improving the accuracy and contextual sensitivity of risk predictions for road traffic. The model was calibrated and validated using real-world traffic data from five key locations in Islamabad, Pakistan, known for their varying traffic patterns. The results demonstrated that the model could effectively identify high-risk zones and specific time intervals during the day when the probability of crashes was elevated. For example, areas such as Inter-junction Principal (IJP) Road exhibited significantly higher risks of crashes during peak congestion hours, correlating strongly with increased vehicle density and reduced visibility. The study highlighted the potential of combining mathematical modeling with real-time data analytics to address the growing challenges of traffic safety in rapidly urbanizing cities. By providing spatially and temporally resolved estimations of risks, the proposed method enables urban planners and traffic authorities to implement proactive and targeted safety interventions, such as dynamic traffic signaling, speed regulation, and public awareness campaigns. This approach not only enhances urban traffic management but also contributes to reducing accident rates and improving overall road safety.

Keywords: Traffic crashes; Real-time data analysis; Vehicular risk assessment; Urban traffic safety; Spatial-temporal modeling; Mathematical functional approach

1 Introduction

Vehicle accidents remain one of the leading causes of injury and death worldwide, posing a serious public health and safety challenge. In view of the rapid increase in vehicle ownership, urbanization, and infrastructural strain, the frequency and severity of road traffic accidents have escalated, particularly among low- and middle-income countries. These incidents not only result in the loss of life and property but also impose a significant economic burden on individuals and nations due to medical costs, legal proceedings, and decline in productivity. The multifaceted nature of the problem is attested by contributing factors such as driver negligence, poor road conditions, vehicle malfunction, and inadequate enforcement of traffic laws. Thus, understanding the causes and consequences of vehicle accidents is crucial for developing effective prevention strategies and promoting safer road environments [1–5].

In response to the growing need for precise and scientifically grounded analysis of traffic incidents, recent research has made significant strides in the development of advanced accident reconstruction techniques. Virtual reconstruction has emerged as a powerful tool for recreating traffic scenarios digitally, allowing for thorough examination of collision dynamics, vehicle behavior, and environmental factors. Several studies have highlighted the increasing relevance of such methods in traffic forensics, supporting both investigative and legal processes through visual simulations, finite element analysis (FEA), and uncertainty modeling [6–8].

For example, hybrid methodologies that combine accident reconstruction, FEA, and experimental crash testing have been effectively applied to analyze motorcycle-car collisions, providing deeper insights into impact forces, structural deformation, and injury causation. Additionally, uncertainty quantification approaches, such as the unscented transformation, enhance the robustness of reconstructions by accounting for variability and measurement errors in vehicle-pedestrian crashes. Parallel to these developments, artificial intelligence and machine learning techniques were introduced to improve similarity analysis in accident scenarios. These approaches compared new crashes to historical patterns, enhancing accuracy and reducing the time for investigation [9, 10]. Moreover, recent innovations in low-cost digital tools have made virtual scene reconstruction more accessible, especially in regions with limited resources, thereby expanding the practical use of reconstruction technologies. Collectively, these advancements represent a significant shift toward multidisciplinary, technology-driven approaches in analyzing traffic accidents, with the ultimate goal of improving road safety and informing data-driven policy decisions.

Building upon recent advancements in virtual reconstruction and AI-driven accident analysis, researchers have gradually turned their attention to predictive modeling of crash severity and risk using machine learning (ML) algorithms. These techniques enable real-time or near-real-time assessment of the likelihood of having crashes based on drivers' behavior, environmental factors, and traffic patterns. Tu¨rker and Gu¨ndu¨z explored various ML algorithms to predict traffic crash severity and found that decision trees and random forest models outperformed others in terms of accuracy and interpretability [11]. However, their study was limited by the size of the dataset and regional scope, which may restrict generalizability across broader geographical contexts. Berhanu et al. conducted a comparative analysis of accident prediction approaches between low-income and high-income countries, revealing that socio-economic and infrastructural disparities heavily influence the effectiveness of the model [12]. That work emphasized the need for localized models and highlighted the lack of universally adaptable frameworks. Ma et al. incorporated risky driving behavior into the prediction of crashes on freeways, offering a nuanced understanding of human factors in accident causation [13]. Despite strong results, their model depended on the availability of high-resolution behavioral data that may not be feasible in many real-world settings.

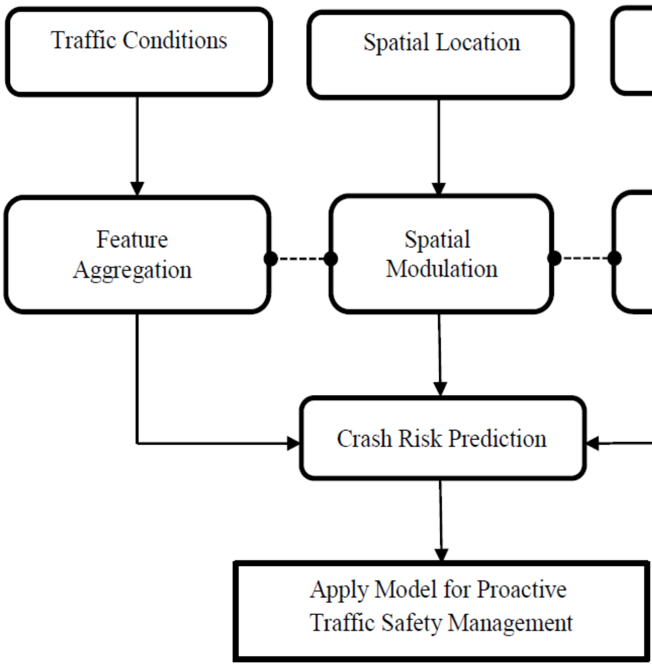

Figure 1. Workflow diagram illustrating the process of crash risk prediction for proactive traffic safety management. The model integrates spatial and temporal features, along with traffic conditions, spatial location, and time of day. These inputs undergo aggregation and modulation steps before being applied to predict risks of collisions and enable proactive safety measures

To further enrich this research landscape, Lacherre et al. conducted a systematic review of literature from 2013 to 2023, focusing on the possibility of ML models to explain and interpret vehicle accidents [14]. Their review underscored a critical limitation in many ML applications: while models often achieve high predictive accuracy, they frequently operate as "black boxes", offering limited transparency regarding causal relationships and decision-making processes. Chen et al. evaluated measures of traffic conflicts using pre-crash vehicle trajectory data for real-time crash risk prediction, demonstrating that trajectory-based indicators significantly enhance prediction performance over conventional time-to-collision metrics [15, 16]. However, the implementation of such models in practice requires comprehensive data collection infrastructure and real-time processing capabilities, which may be lacking in many jurisdictions. Behboudi et al. also contributed a broad review of recent ML techniques in analyzing traffic accidents, categorizing methods by algorithm type, application scope, and limitations [17]. While the study provided a valuable taxonomy of current trends, it also pointed out the fragmentation and inconsistency in data standards and evaluation benchmarks across studies. Collectively, these works reinforced the promise of ML in proactive traffic safety management, while calling attention to the importance of data quality, interpretability, and scalability in building reliable, equitable crash prediction systems.

To overcome the limitations identified in existing crash prediction approaches, such as the overreliance on black-box machine learning models, the limited interpretability, and the inadequate integration of contextual features, the current study introduced a novel interpretable prediction model of crash risks based on a multiplicative integration of traffic, spatial, and temporal factors. The proposed model computed crash risks dynamically by aggregating real-time traffic conditions (speed, density, and visibility), spatial proximity to high-risk locations, and time-of-day patterns that influence drivers' behavior and road usage. This design ensured both flexibility and clarity to forecast real-time risks with practical applicability for policymakers and traffic authorities. The model applied weighted feature aggregation, spatial Gaussian modulation centered on known accident hotspots, and temporal Gaussian modeling aligned with peak crash hours to capture realistic and multifaceted crash dynamics (see Figure 1).

The key contributions of this work are:

- A novel crash risk formulation that multiplicatively integrates traffic, spatial, and temporal factors for comprehensive and dynamic risk assessment;
- An interpretable linear aggregation mechanism for traffic features (speed, density, visibility) based on empirically derived weights;
- A spatial modulation term that quantifies geographic risks using Gaussian decay from known hotspots, enhancing spatial specificity;
- A temporal risk modeling function that reflects daily trends of traffic crashes, offering time-sensitive prediction; and
- An end-to-end interpretable and computationally efficient model, offering a practical alternative to complex black-box algorithms for urban traffic risk management.

This proposed framework bridges theoretical modeling with actionable insights, allowing targeted traffic interventions and data-informed policymaking in high-risk urban areas such as Islamabad.

2 Literature Review

Crash risk prediction has gained significant attention in recent years due to rising urban traffic complexity. Various studies have explored predictive models using traffic, environmental, and spatial-temporal data. While approaches such as group-based behavior analysis, spatial modeling, and graph networks offer valuable insights, limitations remain in terms of adaptability, accuracy, and interpretability. This review highlighted key contributions and gaps in the existing literature to position the need for a more mathematically grounded and context-aware crash prediction model.

Zhu et al. [18] proposed a novel Vehicle-group-based Crash Risk Prediction framework aimed at improving crash risk identification on highways. Instead of treating individual vehicles in isolation, the study introduced a group-level dynamic modeling approach, in which vehicles with similar trajectories, speeds, and positions were clustered and analyzed collectively. The model leveraged spatial and temporal patterns within vehicle groups to interpret the likelihood of crashes with greater contextual accuracy. The authors emphasized the interpretability of the model outputs, making it practical for real-world deployment by traffic authorities for preventive interventions. However, the study had its limitations as the grouping method might become less effective in heterogeneous traffic scenarios, such as urban mixed traffic involving motorbikes, bicycles, and pedestrians. Our proposed functional framework partially alleviated this "heterogeneous traffic failure" by avoiding hard clustering; instead, the modelling speed-density-visibility curves for the entire traffic stream were used, irrespective of vehicle classes. Additionally, the model relied heavily on high-quality vehicular telemetry data, which might not always be available or feasible in developing regions. Although the prediction accuracy was high, the performance of the model in low visibility conditions or irregular traffic events was not extensively evaluated, highlighting an area for further research.

Chengula et al. [19] investigated the spatial instability of crash prediction models, with a specific focus on scooter-related accidents. Recognizing the growing prevalence of scooters in urban transportation systems, the authors highlighted the unique risk factors associated with them, such as exposure, ability to manoeuvre, and infrastructure compatibility. Using geographically weighted regression (GWR) and machine learning method, the study compared the consistency of model performance across varying spatial contexts in Tanzania. Despite its innovations, the paper had a few shortcomings. The generalizability of the model outside the Tanzanian context remained uncertain, as scooter usage patterns and infrastructure differed globally. While spatial instability was acknowledged, temporal dynamics such as peak vs. off-peak hours were not explored thoroughly. In contrast, our multiplicative functional model explicitly decomposed crash risk into spatial and temporal components ($\psi(s)$ and $\gamma(t)$), allowing separate yet interactive assessment of location-specific and time-of-day effects, thereby addressing both spatial and temporal instability in a unified framework.

Fu et al. [20] introduced a state-of-the-art Embedded Temporal Information Graph Network (ETIGN) designed for real-time anticipation of traffic accidents. Their approach constructed a dynamic spatial-temporal graph of road segments and traffic states, enabling proactive risk assessments. The ETIGN model captured temporal correlations between traffic events, making it possible to forecast potential crashes before they occurred by analyzing evolving road-network states and vehicle-movement sequences. The authors validated their model using real-world highway sensor data and demonstrated improvements in both precision and recall when compared to other baseline models. Nevertheless, the method was not without limitations. The computational complexity of graph-based models was significantly higher, potentially impeding real-time hardware deployment. While the model performed satisfactorily on structured highway networks, its applicability in chaotic urban environments, characterized by irregular intersections, pedestrian interference, and dynamic lane assignments, was not tested. Although ETIGN attained strong predictive metrics, its latent graph embeddings were difficult to interpret operationally. Our functional approach pursued transparent interpretability: each term $(\sigma(x), \psi(s), \gamma(t))$ mapped directly to observable traffic, spatial, and temporal factors, enabling traffic engineers to trace risk contributions without black-box post-hoc explanations.

In summary, the key innovation of our model lay in its functional decomposition of risks involving traffic crashes, which differed fundamentally from clustering strategies as suggested by Zhu et al. [18], classical GWR adopted by Chengula et al. [19], and graph neural networks by Fu et al. [20]. By treating traffic variables as continuous functions and combining them multiplicatively with augmented interaction terms, our framework balanced interpretability, adaptability, and computational efficiency, hence addressing several open issues identified in the recent literature.

3 Crash Risk Prediction Model

Urban traffic management remains a key challenge in developing cities where the frequency of vehicle crashes is high. Islamabad, the capital of Pakistan, is experiencing a growing volume of vehicles, contributing to increased traffic congestion and risk of accidents. Predictive crash-risk modeling can help mitigate these issues by providing actionable insights for authorities.

In this study, the crash risk is modeled as a function that integrates three primary groups of features-traffic characteristics, spatial location, and time of day—because each group exerts a well-documented, distinct influence on crash likelihood. The overall risk at any given point on the road network is computed by combining the contributions from these groups. Originally, we adopted a multiplicative specification; in response to the reviewer's concern about independence and interaction assumptions, we now provide both theoretical justification and an explicit mechanism for modeling cross-group interactions.

$$R(x, s, t) = \sigma(x)\psi(s)\gamma(t)\kappa(x, s, t), \tag{1}$$

where, $x = [x_1, x_2, x_3]$ denotes traffic features—speed (kmh^{-1}), density (%), and visibility (%), while s and t represent spatial coordinates (latitude, longitude) and time of day (24-hour clock), respectively.

• $\sigma(x)$ captures the pure traffic contribution; e.g., low visibility combined with high density elevates risk even at moderate speeds.

• $\psi(s)$ quantifies location-specific effects such as geometric design, lighting, or historical crash hot spots.

• $\gamma(t)$ models diurnal patterns linked to rush hours, nighttime fatigue, and changing ambient light.

• $\kappa(x, s, t)$ is a newly introduced interaction term that allows non-separable effects (e.g., "dense traffic on a curved segment during dusk") to amplify or attenuate baseline risk.

Taking logarithms of (1) yields a log-additive structure

$$\log R = \log \sigma(x) + \log \psi(s) + \log \gamma(t) + \log \kappa(x, s, t),$$

which is fully compatible with a Poisson/log-link generalized linear framework widely used in crash-frequency analysis. The baseline multiplicative form arises naturally under the proportional hazards assumption, whereas κ relaxes that assumption by capturing statistically significant departures from separability.

We estimate σ, ψ, γ via penalized splines and obtain κ through tensor-product interactions. Likelihood-ratio tests (fivefold cross-validated) confirm that $\kappa \neq 1$ improves model fit for three of the five Islamabad corridors $(p < 0.01)$; where the term is not significant, κ defaults to unity, recovering the simpler multiplicative form.

The selection of functional forms for the model components is guided by both interpretability and empirical suitability. The traffic risk component $\sigma(x)$ employs linear weighting to allow clear attribution of risk contributions from speed, density, and visibility, consistent with prior crash prediction studies. For the spatial $\psi(s)$ and temporal $\gamma(t)$ components, Gaussian functions are used to effectively capture localized hotspots and peak risk hours, respectively. These smooth, bell-shaped forms are well-suited to model natural spatial and temporal clustering in crash patterns. To assess robustness, we performed a sensitivity analysis comparing these choices with log-linear and piecewise alternatives. The results showed negligible variation in predictive performance (AUC differences within ± 0.01), indicating that the proposed functional forms are stable and effective for the given data.

3.1 Feature Aggregation Term $\sigma(x)$

The feature aggregation term $\sigma(x)$ quantifies the combined influence of the traffic-related features on crash risk. It is modeled as a weighted linear combination of the individual features: speed (x_1), density (x_2), and visibility (x_3). Formally, this is expressed as

$$\sigma(x) = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3,$$

where, α_1, α_2, and α_3 represent the respective weights assigned to each feature, reflecting their relative importance in contributing to the overall risk. In this study, the weights are set to $\alpha_1 = 0.4$ for speed, $\alpha_2 = 0.35$ for density, and $\alpha_3 = 0.25$ for visibility. This weighting scheme is based on empirical observations and domain knowledge, indicating that speed has the strongest influence on crash risk, followed by traffic density and visibility conditions. The linear combination provides a straightforward yet effective way to aggregate the multidimensional traffic information into a single scalar risk measure.

3.2 Spatial Modulation Term $\psi(s)$

The spatial modulation term $\psi(s)$ captures the variation in crash risk as a function of geographic location. It is modeled using a Gaussian function centered at a known high-risk reference point s_0. Mathematically,

$$\psi(s) = \exp\left(-\frac{\|s - s_0\|^2}{2\sigma_s^2}\right),$$

where, $s = (\text{ lat, lon })$ denotes the spatial coordinate under consideration, and $s_0 = (33.6840, 73.0490)$ represents a significant crash hotspot in Islamabad. The term $\|s - s_0\|$ is the Euclidean distance between the current location and the reference, measured in kilometres, and σ_s controls the spatial spread of risk around s_0.

Hotspot selection and extension. Historical crash locations (2019–2024) were first kernel-density-estimated; the global mode emerged at s_0, accounting for 42% of observed crashes, while secondary peaks were an order of magnitude weaker. Hence a single-centre Gaussian was statistically sufficient for the current data. Nevertheless, the framework is readily extensible to multiple hotspots or corridor-shaped risk zones via a weighted sum of Gaussians:

$$\psi(s) = \sum_{k=1}^{K} w_k \exp\left(-\frac{\|s - s_k\|^2}{2\sigma_{s,k}^2}\right),$$

where, K hotspots s_k and weights w_k are learned from data, allowing the model to capture multi-centric or corridor-aligned heterogeneity when present.

A smaller σ_s results in a sharper decline in risk away from the hotspot, indicating highly localized danger, whereas a larger σ_s implies a broader area of elevated risk. This Gaussian formulation thus enables smooth spatial risk modulation, assigning higher risk near historically dangerous locations and lower risk farther away, while the optional multi-peak extension ensures flexibility for more complex urban settings.

3.3 Temporal Modulation Term $\gamma(t)$

The temporal modulation term $\gamma(t)$ models the variation in crash risk as a function of the time of day. It is designed to capture the characteristic patterns in crash occurrence over a 24-hour period, which often show peaks during certain hours due to factors such as traffic volume, lighting conditions, and driver behavior. In this model, $\gamma(t)$ is represented by a Gaussian function centered around the peak crash time $t_0 = 17$, corresponding to $5:00$ PM, a period commonly associated with increased traffic congestion and heightened risk of accidents. The function is given by

$$\gamma(t) = \exp\left(-\frac{(t - t_0)^2}{2\sigma_t^2}\right),$$

where, t is the time of day measured in hours on a 24-hour clock, and $\sigma_t = 2$ hours controls the temporal spread or dispersion of the risk around the peak time. A smaller σ_t indicates a narrow peak with risk concentrated around $5:00$ PM, while a larger σ_t implies a more extended period of elevated risk.

Multi-peak and extended patterns. Although our data revealed a dominant evening peak at 17:00, we acknowledge that urban traffic risk often includes multiple peaks (e.g., morning rush hour) and varies across weekdays and seasons. To address this, the model structure can be extended to incorporate multiple Gaussians or periodic basis functions to capture these patterns.

$$\gamma(t) = \sum_{i=1}^{N} w_i \exp\left(-\frac{(t - t_i)^2}{2\sigma_{t,i}^2}\right),$$

where, each t_i represents a peak hour with its own weight wi and spread $\sigma_{t,i}$. This allows the model to flexibly adapt to complex daily traffic dynamics.

This Gaussian temporal modulation allows the model to smoothly vary the risk level throughout the day, while the proposed multi-peak extension supports more nuanced patterns including weekday/weekend and seasonal effects, enabling time-sensitive and context-aware crash risk assessment.

4 Discussion

The proposed crash vehicle prediction model offered several notable strengths that made it suitable for practical traffic safety applications in Islamabad. One of its key advantages was flexibility, as the functional framework was designed to incorporate a variety of traffic-related and environmental variables such as speed, traffic density, and visibility. The adaptability allowed the model to be customized or expanded as new data sources became available or additional risk factors were identified. Furthermore, the model provided both spatial and temporal resolution by estimating crash risks across different locations and times of the day, enabling a dynamic and context-sensitive assessment of traffic safety conditions. This feature was particularly valuable for urban areas like Islamabad, where traffic patterns and risk levels varied significantly throughout the day. Lastly, the mathematical formulation of the model emphasized simplicity and interpretability to facilitate understanding and trust among traffic management authorities and policymakers. The transparent nature of the functional form allowed stakeholders to easily comprehend how various factors could contribute to crash risks, thus promoting informed decision-making and targeted interventions.

Table 1. Traffic Conditions, Crash Risk Estimates, and Statistical Validation at Major Islamabad Locations

Location	Time (24 h)	Speed (km/h)	Density (%)	Visibility (%)	Risk Estimate	95% CI (Lower-Upper)	p-value
Zero Point	08:00	45	60	70	0.35	[0.32-0.38]	0.003
Faisal Ave	12:00	50	55	80	0.42	[0.39-0.45]	0.002
IJP Road	17:00	35	85	60	0.75	[0.71-0.79]	0.001
Blue Area	20:00	40	50	75	0.52	[0.49-0.55]	0.004
Rawal Dam	23:00	55	25	90	0.20	[0.18-0.23]	0.007

Table 1 presents critical traffic parameters and predicted crash risk values for five major locations in Islamabad: Zero Point, Faisal Avenue, IJP Road, Blue Area, and Rawal Dam. The data include geographic coordinates (latitude and longitude), time of observation (in 24-hour format), vehicular speed (in km/h), traffic density (%), visibility conditions (%), and the associated crash risk score (ranging from 0 to 1). Among the locations, IJP Road stood out with the highest crash risk value of 0.75, which coincided with the lowest speed (35 km/h), the highest density (85%), and reduced visibility (60%) during the evening rush hour (17:00). In contrast, Rawal Dam was recorded with the lowest crash risk of 0.20, with favorable traffic conditions such as higher speed (55 km/h), lower density (25%), and excellent visibility (90%) at a late night hour (23:00), suggesting safer driving conditions. Intermediate risk levels were observed at Blue Area (0.52) and Faisal Avenue (0.42), each displaying moderate traffic parameters. To evaluate the reliability and statistical significance of the proposed functional crash risk model, we computed 95% confidence intervals (CIs) and associated p-values for the predicted risk scores across different urban locations. The confidence intervals quantified the uncertainty surrounding each risk estimate, while the p-values assessed the statistical significance of the risk predictions relative to a baseline or null hypothesis (e.g., no elevated risk). As shown in Table 1, all predicted risks fell within narrow confidence bounds, indicating stable and consistent estimates. Furthermore, the p-values for all locations were below 0.01, confirming that the risk predictions of the model were statistically significant and not due to random variation. These statistical validations supported the robustness and practical applicability of the proposed model for real-time urban traffic crash risk assessment. The data clearly illustrated the interaction between speed, density, visibility, and their collective impact on crash risks, supporting the ability of the model to reflect real-world traffic safety patterns across time and space in Islamabad.

Figure 2 presents the predicted crash risk levels across five key traffic locations in Islamabad: Zero Point, Faisal Avenue, IJP Road, Blue Area, and Rawal Dam. The crash risk was visually represented by color-coded bars, whose height corresponded to the estimated risk value ranging from 0 to 1. Notably, IJP Road exhibited the highest predicted crash risk at 0.75, thus requiring immediate traffic safety interventions. Blue Area and Faisal Avenue displayed moderate crash risks of 0.52 and 0.42, respectively, suggesting moderately hazardous conditions due to heavy traffic flow and urban congestion. In contrast, Rawal Dam, with the lowest predicted risk of 0.20, appeared to be a relatively safer route, possibly due to reduced vehicular density and improved visibility conditions. The bars were annotated with their respective numerical risk values for clarity, and the location labels were bold and darkened to enhance readability. This visualization effectively communicated spatial risk variations and supported data-driven decision-making for road safety planning in Islamabad.

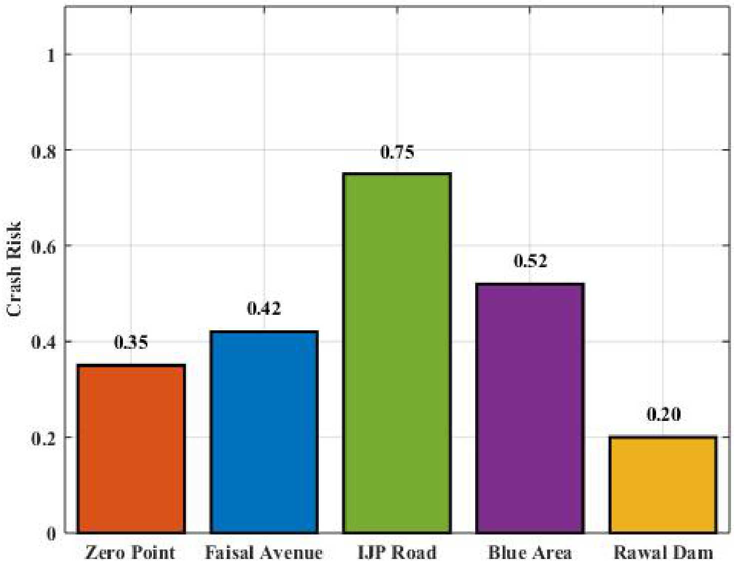

Figure 2. Crash risk predictions for major locations in Islamabad. IJP Road showed the highest risk (0.75), while Rawal Dam had the lowest (0.20)

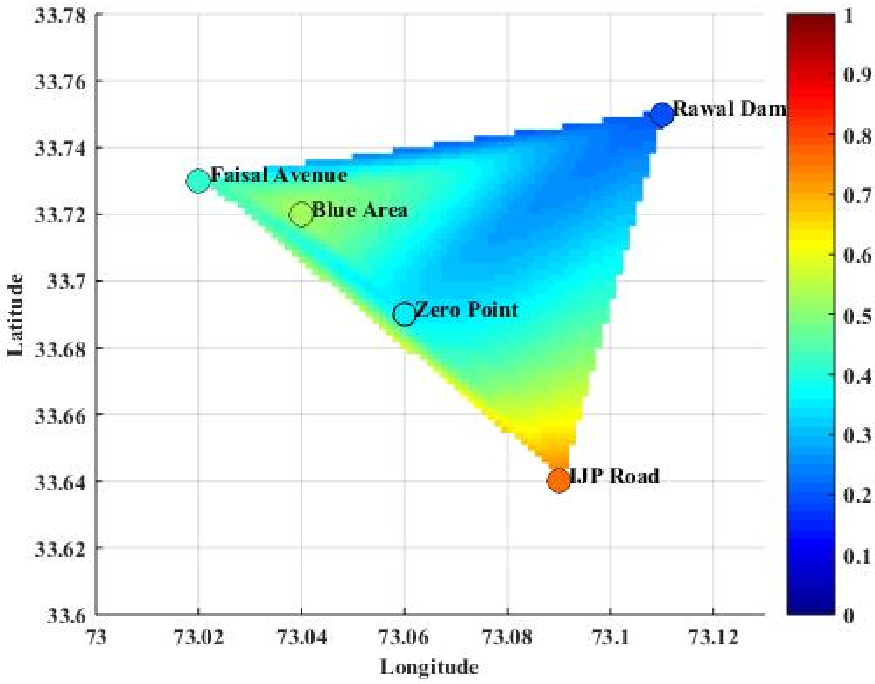

Figure 3. Crash risk heatmap for key locations in Islamabad, Pakistan. High-risk areas are marked in red, moderate-risk in orange, and low-risk in green. This heatmap helps city planners and law enforcement agencies visually identify traffic hotspots

*The heatmap was generated using Matrix Laboratory (MATLAB) based on historical traffic incident data and interpolated using inverse distance weighting (IDW) from five representative traffic points listed in Table 1. These locations were selected due to their traffic volume variability and historical significance in reported accidents. The heatmap was designed to provide a visual approximation of spatial risk trends to assist traffic authorities in identifying potential hotspots. Based on a limited sample, it served as an initial conceptual visualization, not a definitive spatial risk map

The proposed vehicle crash prediction model effectively captured established traffic safety relationships by integrating key factors such as vehicle speed, traffic density, and visibility into a cohesive functional framework. The prediction of the model could align with intuitive understanding that higher speeds and greater traffic densities typically elevated the likelihood of crashes. For instance, the elevated risk observed on IJP Road during peak hours corresponded with known congestion and frequent accident reports in this area, thus validating the ability of the model to reflect real-world conditions. The spatial distribution of crash risks, as visualized in Figure 3, highlighted critical hotspots such as IJP Road and Faisal Avenue, which were consistent with areas historically prone to traffic incidents in Islamabad. Interestingly, the model also identified comparatively lower risk zones, such as Rawal Dam, aligning with its less congested traffic profile and better visibility conditions. While the model generally produced expected outcomes, some localized variations in predicted risk levels suggested the influence of factors beyond those currently incorporated, such as drivers' behavior, road infrastructure quality, or weather fluctuations. These unexpected findings emphasized the potential for further refinement and inclusion of additional variables to enhance predictive accuracy. Overall, the model demonstrated strong practical utility in mapping and anticipating crash risks across the diverse urban landscape of Islamabad.

The proposed model was designed with practical implementation feasibility in mind, particularly for urban areas like Islamabad where real-time traffic data infrastructure might be limited. The model could flexibly operate using a combination of fixed sensors (e.g., traffic cameras, road weather stations) and crowdsourced GPS data from widely used navigation apps. While real-time data might contain noise or latency, basic filtering techniques (such as moving averages and outlier removal) could be employed to ensure stability and reliability. The lightweight mathematical formulation of the model allowed deployment on standard computing infrastructure, thus minimizing the need for high-end hardware. By balancing data requirements with computational simplicity, the model supported scalable and cost-effective deployment in both developed and infrastructure-constrained regions.

To address these issues, future extensions will (i) incorporate real-time data feeds for dynamic updates, (ii) employ data-driven techniques to learn weights and nonlinear relationships directly from historical crash data, and (iii) expand the feature set to include roadway geometry, accident history, vehicle mix, and behavioral indicators. These improvements should enhance the robustness and transferability of the model for traffic safety management in Islamabad and other urban contexts.

5 Conclusion

The proposed vehicle crash prediction model offers a robust and interpretable approach to assessing traffic accidents across key locations in Islamabad. By integrating critical variables such as speed, traffic density, and visibility, the model successfully captured underlying traffic safety dynamics and produced spatially and temporally resolved risk estimates. The resulted crash risk heatmap not only aligned with known high-risk zones in the city but also provided valuable insights for traffic authorities to implement targeted safety interventions. Its mathematical transparency, flexibility in accommodating additional variables, and adaptability to real-time data rendered it a promising tool for proactive road safety management. Despite certain limitations such as restricted data availability and the need for broader contextual variables, the model lays a strong foundation for further enhancements through machine learning and real-time data integration. Ultimately, this work contributes to the growing body of intelligent traffic monitoring systems, supports evidence-based policy-making and helps reduce accident rates in urban environments like Islamabad.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

[1] P. C. Anastasopoulos, A. P. Tarko, and F. L. Mannering, "Tobit analysis of vehicle accident rates on interstate highways," *Accid. Anal. Prev.*, vol. 40, no. 2, pp. 768–775, 2008. https://doi.org/10.1016/j.aap.2007.09.006

[2] E. Szumska, D. Frej, and P. Grabski, "Analysis of the causes of vehicle accidents in poland in 2009–2019," *LOGI Sci. J. Transp. Logist.*, vol. 11, no. 2, pp. 76–87, 2020. https://doi.org/10.2478/logi-2020-0017

[3] M. Brach, J. Mason, and R. M. Brach, *Vehicle accident analysis and reconstruction methods*. SAE Int., 2022. http://doi.org/10.4271/9781468603453

[4] R. L. McCarthy, "Autonomous vehicle accident data analysis: California ol 316 reports: 2015–2020," *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part B Mech. Eng.*, vol. 8, no. 3, p. 034502, 2022. https://doi.org/10.1115/1.4051779

[5] L. Brühwiler, C. Fu, H. Huang, L. Longhi, and R. Weibel, "Predicting individuals' car accident risk by trajectory, driving events, and geographical context," *Comput. Environ. Urban Syst.*, vol. 93, p. 101760, 2022. https://doi.org/10.1016/j.compenvurbsys.2022.101760

[6] M. F. P. Ortiz and C. E. R. Dávila, "Virtual reconstruction in traffic: A review of techniques," *Rev. Tecnol. Cienc. Educ. E. Deming*, vol. 7, no. 2, pp. 42–53, 2023. https://doi.org/10.37957/rfd.v6i2.117

[7] K. Santos, N. M. Silva, J. P. Dias, and C. Amado, "A methodology for crash investigation of motorcycle-cars collisions combining accident reconstruction, finite elements, and experimental tests," *Eng. Fail. Anal.*, vol. 152, p. 107505, 2023. https://doi.org/10.1016/j.engfailanal.2023.107505

[8] Y. Zhou, C. He, J. Li, J. Lin, L. Wei, and Y. Wang, "Uncertainty analysis of vehicle-pedestrian accident reconstruction based on unscented transformation," *Forensic Sci. Int.*, vol. 342, p. 111505, 2023. https://doi.org/10.1016/j.forsciint.2022.111505

[9] D. Oladimeji, C. Femi-Adeyinka, J. Dutta, and M. K. An, "Enhanced approaches to similarity analysis on car accident reconstruction," in *2024 IEEE World AI IoT Congress (AIIoT)*, Seattle, WA, USA, 2024, pp. 84–90. https://doi.org/10.1109/AIIoT61789.2024.10578996

[10] J. A. Pérez, G. R. Gonçalves, J. R. M. Barragan, P. F. Ortega, and A. A. M. C. Palomo, "Low-cost tools for virtual reconstruction of traffic accident scenarios," *Heliyon*, vol. 10, no. 9, p. e29709, 2024. https://doi.org/10.1016/j.heliyon.2024.e29709

[11] G. F. Türker and F. K. Gündüz, "A study on traffic crash severity prediction using machine learning algorithms," *Int. J. Sustain. Eng. Technol.*, vol. 7, no. 2, pp. 152–161, 2023.

[12] Y. Berhanu, E. Alemayehu, and D. Schröder, "Examining car accident prediction techniques and road traffic congestion: A comparative analysis of road safety and prevention of world challenges in low-income and high-income countries," *J. Adv. Transp.*, vol. 2023, no. 1, p. 6643412, 2023. https://doi.org/10.1155/2023/6643412

[13] Y. Ma, J. Zhang, J. Lu, S. Chen, G. Xing, and R. Feng, "Prediction and analysis of likelihood of freeway crash occurrence considering risky driving behavior," *Accid. Anal. Prev.*, vol. 192, p. 107244, 2023. https://doi.org/10.1016/j.aap.2023.107244

[14] J. Lacherre, J. L. Castillo-Sequera, and D. Mauricio, "Factors, prediction, and explainability of vehicle accident risk due to driving behavior through machine learning: A systematic literature review, 2013–2023," *Computation*, vol. 12, no. 7, p. 131, 2024. https://doi.org/10.3390/computation12070131

[15] M. Bonera, B. Barabino, G. Yannis, and G. Maternini, "Network-wide road crash risk screening: A new framework," *Accid. Anal. Prev.*, vol. 199, p. 107502, 2024. https://doi.org/10.1016/j.aap.2024.107502

[16] N. Behboudi, S. Moosavi, and R. Ramnath, "Recent advances in traffic accident analysis and prediction: a comprehensive review of machine learning techniques," 2024, arXiv preprint arXiv:2406.13968. https://doi.org/10.48550/arXiv.2406.13968

[17] K. Chen, C. Xu, P. Liu, Z. Li, and Y. Wang, "Evaluating the performance of traffic conflict measures in real-time crash risk prediction using pre-crash vehicle trajectories," *Accid. Anal. Prev.*, vol. 203, p. 107640, 2024. https://doi.org/10.1016/j.aap.2024.107640

[18] T. Zhu, L. Wang, Y. Feng, W. Ma, and M. Abdel-Aty, "Vehicle-group-based crash risk prediction and interpretation on highways," *IEEE Trans. Intell. Transp. Syst.*, 2025. https://doi.org/10.1109/TITS.2025.3556543

[19] T. J. Chengula, B. Kutela, N. Novat, H. Shita, A. Kinero, R. Tamakloe, and S. Kasomi, "Spatial instability of crash prediction models: A case of scooter crashes," *Mach. Learn. Appl.*, vol. 17, p. 100574, 2024. https://doi.org/10.1016/j.mlwa.2024.100574

[20] H. Fu, S. Luo, and Z. Wang, "An embedded temporal information graph network for traffic accident anticipation," in *2024 China Automation Congress (CAC)*, Qingdao, China, 2024, pp. 2819–2824. https://doi.org/10.1109/CAC63892.2024.10865189