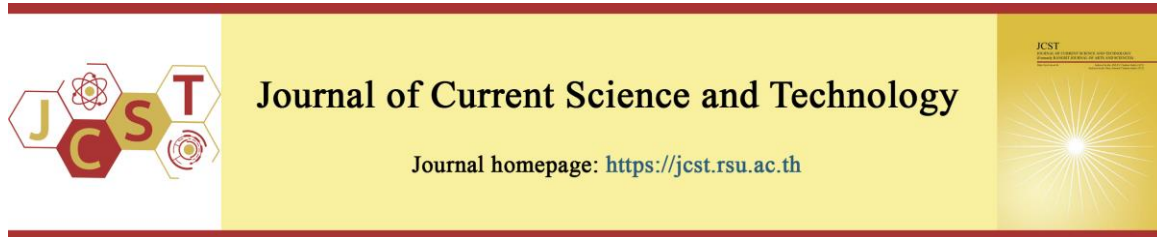


Cite this article: Vardhan, R., & Ram, S. (2026). Integrating perceived risk and spatial data for urban junction safety: a micro-spatial machine learning approach. *Journal of Current Science and Technology*, 16(2), Article 185. <https://doi.org/10.59796/jcst.V16N2.2026.185>



Integrating Perceived Risk and Spatial Data for Urban Junction Safety: A Micro-Spatial Machine Learning Approach

Rohan Vardhan*, and Sewa Ram

Department of Transport Planning, School of Planning and Architecture, New Delhi, India

*Corresponding author; E-mail: rohan.phd394tp23@spa.ac.in

Received 27 November 2025; Revised 28 December 2025; Accepted 22 January 2026; Published online 7 April 2026

Abstract

Road traffic crashes (RTCs) remain a leading global cause of mortality, yet conventional safety analyses often overlook how individuals perceive road risk, a key behavioral dimension influencing crash likelihood. In many developing cities, the absence of comprehensive crash databases further limits evidence-based safety planning. This study addresses these challenges by developing a transferable ensemble machine learning framework across junctions that predicts crash-risk hotspots using perceived risk as a proxy under data-scarce conditions. The framework integrates GIS (Geographic Information Systems)-based infrastructure mapping, KDE (Kernel Density Estimation)-derived spatial perception maps, PROMETHEE-based parameter prioritization, and an ensemble of Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Decision Trees (DT) to enable scalable, microspatial safety analysis. Perceptions of risk across behavioral, infrastructure, and environmental factors were systematically mapped using Geographic Information Systems (GIS) and modeled through the ensemble model. The optimized ensemble achieved high predictive performance ($R^2 = 0.92$, $\kappa = 0.779$), effectively integrating psychological perception with spatial infrastructure data at a micro-spatial resolution of $1 \text{ m} \times 1 \text{ m}$. Strong correlations emerged between perceived risk and determinants such as tailgating tendency ($r = 0.9347$), building density and land-use mix ($r = 0.9497$), and poor lighting in dense areas ($r = -0.3714$). The resulting Crash Risk Perception Index (CRPI) revealed distinct spatial clusters, distinguishing validated danger zones from perceptual-empirical divergences. Hidden hazard zones (6.1% of cells) - high crash frequency but low perceived risk - and near-miss zones (12.7% of cells) - low crash frequency but high perceived risk - were identified as priority locations for targeted awareness, enforcement, and design interventions. Validation across three additional urban junctions ($\kappa = 0.72\text{--}0.83$) demonstrated robustness and transferability. The framework enables cities lacking crash databases to proactively identify high-risk and near-miss zones, supporting perception-informed and evidence-based strategies for safer, more resilient urban mobility systems.

Keywords: crash risk perception; ensemble machine learning; geographic information systems; near-miss identification; road safety; spatial analysis; urban mobility

1. Introduction

Road traffic crashes (RTCs) claim 1.3 million lives and cause 50 million injuries annually, making them a leading cause of death among young people worldwide (World Health Organization, 2021). Despite global efforts, the UN Sustainable Development Goal 3.6 target to halve RTC deaths by 2020 was not achieved, prompting renewed commitments for

2021–2030. RTCs impose economic losses of approximately 3% of GDP in most nations, with over 90% of fatalities concentrated in low- and middle-income countries. Urban junctions - where vehicle streams, pedestrian flows, and cyclists converge under variable signals, visibility, and behavioral conditions - represent disproportionate concentrations of crash risk, making them the primary focus for evidence-

based safety intervention (Jadaan et al., 2018; Wang et al., 2021a; Choudhary et al., 2015).

Conventional approaches rely on police-reported crash records that are inherently retrospective and, in many developing cities, capture as few as 30–50% of actual injury events (Abdulhafedh, 2017). Perceived crash risk - the subjective assessment of crash likelihood and consequence (Sjöberg et al., 2004) - offers a behaviorally grounded and spatially collectible complement. Research confirms that perceived risk is shaped by infrastructure cues such as lighting, visibility, and lane configuration, and that perception–reality mismatches directly influence speed choice, gap acceptance, and crossing behavior (Lund & Rundmo, 2009; Andersson, 2011; Bucsuházy et al., 2020; Ram & Chand, 2016; Zhang et al., 2019). GIS-based spatial analysis and ensemble machine learning have further demonstrated strong potential for hotspot identification and crash prediction under data-scarce conditions (Keken et al., 2019; Ma et al., 2020; Berhanu et al., 2023; Prasannakumar et al., 2011; Yao et al., 2016; Sajed et al., 2019).

Despite these advances, and building on the extensive examination of crash risk perception and human factors in the spatial realm by Shahi et al. (2023), critical gaps remain. GIS-based crash studies identify hotspots but miss behavioral drivers, while perception surveys capture psychological factors without spatial or infrastructure precision (de Blaeij & van Vuuren, 2003). No validated framework integrates micro-spatial perceived risk with infrastructure attributes using ensemble machine learning for predictive mapping. Moreover, the divergence between perceived and recorded risk - “hidden hazard” and “near-miss” zones - remains underexplored, limiting targeted interventions, especially in data-scarce cities (Cabrera-Arnau et al., 2020).

To address these gaps, this study develops a transferable, perception-driven ensemble machine learning framework demonstrated in Sonipat, India - a rapidly growing National Capital Region city with limited crash records. The framework integrates GIS infrastructure mapping, KDE-derived perception surfaces from participatory surveys, PROMETHEE-based parameter prioritization, and an optimized ANN–SVM–DT ensemble producing crash risk predictions at 1 m × 1 m resolution. By treating perceived risk as both a diagnostic lens and a predictive proxy, it identifies validated danger zones, hidden hazards, and near-miss clusters without reliance on complete crash databases. Validation across three junctions confirms transferability,

providing a generalizable tool for evidence-based urban safety planning.

1.1 Literature Review

Road traffic crash (RTC) research has evolved along two largely parallel streams - spatial-statistical analysis and risk perception - that rarely converge into an integrated predictive framework. Understanding where each stream succeeds and where it falls short is essential for identifying the gap this study addresses.

Spatial analysis of RTCs has established that crashes are not randomly distributed but form statistically significant clusters or hotspots concentrated at specific road segments and junctions (Black, 1991; Levine et al., 1995; Steenberghen et al., 2004; Kam, 2003; Petch & Henson, 2000; Xie & Yan, 2013; Soltani & Askari, 2017). Geographic Information Systems (GIS) have become the primary instrument for visualizing and analyzing these patterns, supporting applications from basic crash geocoding to advanced network-based kernel density estimation (KDE) and spatial autocorrelation modeling (Prasannakumar et al., 2011; Yao et al., 2016; Keken et al., 2019). Grid-based cluster analysis has refined spatial precision, enabling identification of micro-level risk concentrations within junctions (Steenberghen et al., 2004) while spatial autocorrelation remains a critical methodological concern in cluster detection (Black & Thomas, 1998). Built-environment attributes - street width, number of lanes, lighting, land-use density, and proximity to pedestrian facilities - have been consistently linked to crash frequency and severity (Dai, 2012; Rankavat & Tiwari, 2016; Retting et al., 2003; Brüde & Larsson, 2000). Hotspot identification further supports the optimization of safety investment and promotes evidence-based infrastructure improvements (Sajed et al., 2019). However, spatial approaches rely on historical crash data, making them retrospective and unable to explain behavioral mechanisms or function in data-scarce contexts (Abdulhafedh, 2017).

The second stream of research centers on human factors and risk perception. RTCs are recognized as products of complex interactions among driver behavior, vehicle characteristics, and the road environment, with human factors implicated in most of the crash causation (Bucsuházy et al., 2020; Petridou & Moustaki, 2000). Behavioral contributors include speeding, distraction, tailgating, and risky maneuvers (Hallett et al., 2012; Özkan et al., 2012; Albert & Bekhor, 2019; Adanu et al., 2017; Akanbi et al., 2009), all of which are modulated by psychological states such as stress, overconfidence, and thrill-seeking (Elander et al., 1993; Legree et al., 2003;

McKenna et al., 2006; Machado-León et al., 2016). Critically, individuals tend to underestimate their own crash risk while attributing danger to external factors, a bias that sustains unsafe behavior (Andersson, 2011). Perceived crash risk - defined as a road user's subjective assessment of crash likelihood and consequence (Sjöberg et al., 2004) - varies systematically by transport mode, demographic characteristics, and route familiarity: car drivers perceive lower risk than non-motorized users, and experienced commuters often habituate to hazards that occasional users find threatening (Lund & Rundmo, 2009; Ram & Chand, 2016). Environmental and infrastructural cues directly shape these perceptions; lighting, visibility, lane configuration, pedestrian infrastructure, and land-use density have all been linked to perceived safety levels (Bivina & Parida, 2019; Rankavat & Tiwari, 2016; von Stülpnagel & Lucas, 2020; Wang et al., 2021b; Vergel-Tovar et al., 2020; Iamtrakul et al., 2023; Ewing & Dumbaugh, 2009). Despite this evidence, risk perception data are rarely incorporated into spatial safety planning, leaving a critical behavioral dimension unexploited in hotspot identification and intervention design (Guerrero et al., 2020; Shahi et al., 2023; de Blaeij & van Vuuren, 2003).

Machine learning has emerged as a powerful complement to GIS-based spatial analysis for crash prediction. Ensemble methods integrating Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Decision Trees (DT) consistently outperform individual models by capturing non-linear relationships among high-dimensional predictor sets while improving generalization and reducing overfitting (Ma et al., 2020; Berhanu et al., 2023; Halabaku & Bytyçi, 2024). ANN models excel at learning complex behavioral patterns; SVM provides robustness against overfitting through margin-based classification; and DT contributes interpretability through explicit variable importance rankings (Panimalar & Krishnakumar, 2023). When integrated with GIS, these methods enable spatially explicit risk prediction at resolutions previously impractical with statistical regression alone (Xiao et al., 2021). However, existing ML applications in road safety have overwhelmingly relied on objective crash records as the response variable. The use of perceived risk as both a predictor and a response variable within a spatially explicit ensemble framework - particularly at micro-spatial resolution - has not been previously demonstrated.

The literature highlights a key gap: no validated framework integrates perceived crash risk

with micro-spatial infrastructure data using ensemble machine learning for junction-level safety prediction (ZahranEl-Said et al., 2019; Berhanu et al., 2023). Existing studies treat crashes and perception separately at coarse scales rather than modeling their relationship (Mukherjee & Mitra, 2020; Shahi et al., 2023). The divergence between perceived and recorded risk - "hidden hazard" and "near-miss" zones - remains unmapped, limiting targeted interventions. This study addresses these gaps by proposing a transferable, perception-driven micro-spatial framework for predictive urban junction safety analysis in data-scarce contexts.

2. Objectives

This study develops a novel, transferable ensemble machine learning framework for micro-spatial crash risk prediction at urban junctions, addressing the critical gap between subjective risk perception and objective crash analysis in data-scarce environments. The specific objectives are to: (i) systematically map perceived crash risk at $1\text{ m} \times 1\text{ m}$ resolution using GIS-based participatory surveys and Kernel Density Estimation; (ii) identify and prioritize the infrastructure, behavioral, and environmental parameters most strongly associated with perceived crash risk through PROMETHEE-based multi-criteria analysis; (iii) develop and optimize an ensemble machine learning model integrating Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Decision Trees (DT) to predict spatial crash risk from perceived risk inputs; (iv) validate the model's predictive reliability by correlating predicted risk surfaces with recorded crash data and assessing spatial agreement using the kappa coefficient; and (v) evaluate the framework's transferability across four urban junctions with varying geometry, land use, and traffic composition.

2.1 Research Questions

The study is guided by four research questions, each addressed through a corresponding phase of the five-phase methodology described in Section 3:

RQ1: Which infrastructure features strongly influence perceived crash risk at the micro-spatial level?

RQ2: Can perceived risk reliably predict crash hotspots using only infrastructure data?

RQ3: Where do perceived and actual crash risks diverge, and what do these mismatches imply for safety planning?

RQ4: Is the developed model transferable across different urban junctions?

These questions are addressed sequentially across the study phases. RQ1 is tackled in Phase 2, where systematic perception surveys and GIS infrastructure mapping jointly establish which physical and behavioral attributes most strongly covary with perceived danger at the micro-spatial level. RQ2 is addressed in Phases 3 and 4, where the spatial integration of survey-derived risk surfaces with infrastructure variables forms the input dataset for the ensemble machine learning model, enabling a direct test of whether infrastructure data alone can reliably predict crash risk hotspots without reliance on recorded crash histories. RQ3 is resolved in Phase 4 through spatial overlay analysis, which compares model-predicted risk surfaces against recorded crash data to delineate convergent zones - where perception and crashes align - as well as hidden hazard zones of underestimated risk and near-miss zones of overestimated risk, each requiring distinct intervention strategies. RQ4 is examined in Phase 5, where the trained model is applied without recalibration to three additional junctions with differing geometry, land use, and traffic composition, with performance evaluated using the kappa coefficient and overall accuracy to confirm transferability across diverse urban contexts.

3. Materials and Methods

The study followed a five-phase process (Figure 1): Phase 1 covers study area selection/delineation and crash data collection, Phase 2 conducts risk perception surveys and infrastructure mapping, Phase 3 involves data cleaning and spatial integration of all datasets, Phase 4 develops the ensemble machine learning model for crash risk prediction, and Phase 5 validates the results through statistical and spatial correlations among perceived, predicted, and recorded crash data to derive actionable safety insights.

3.1 Study Area and Crash Data (Phase-1)

Sonipat, a planned industrial city in Haryana, India, and part of the National Capital Region (NCR), served as the study area. Located 44 km from New Delhi, its rapid growth, contrasted with stagnant infrastructure, makes it suitable for examining evolving urban road-user perceptions in emerging cities. The city's busiest road, characterized by diverse land use and complex traffic, was selected as the study site (Verma et al., 2022). This junction was chosen for its representativeness of urban conditions and its ability to capture variations in risk perception.

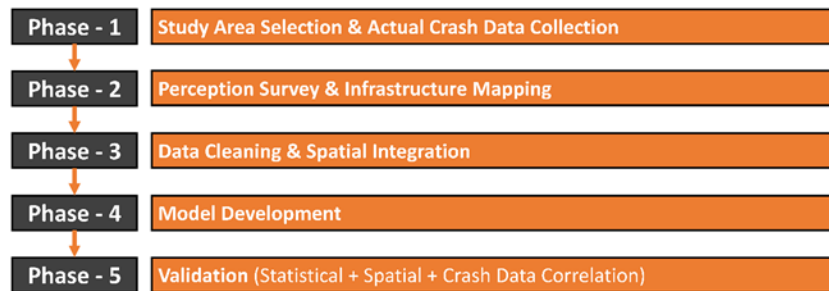


Figure 1 Five-phase methodology framework for micro-spatial crash risk prediction.

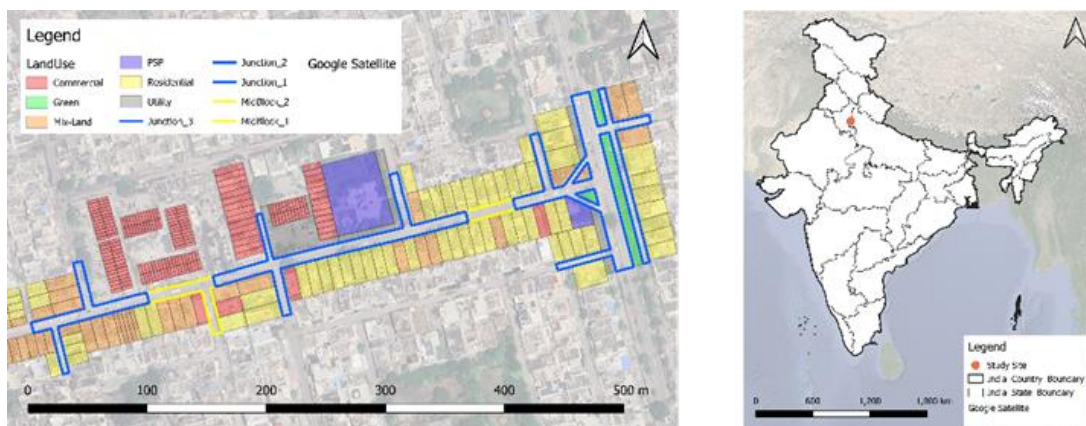


Figure 2 Study area map showing land-use zones, junction boundaries, and geographic location within India (Sonipat, Haryana, NCR).

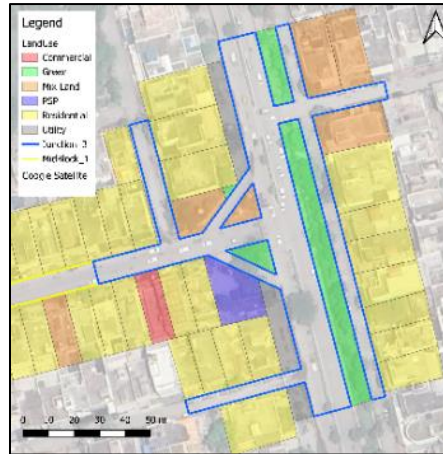


Figure 3 Aerial view of the primary study junction in Sonipat, Haryana, India, showing road geometry, land-use context, and traffic environment.



Figure 4 Aerial views of the three validation junctions: V1 (left) — low-density, mixed land use; V2 (centre) — similar urban context, different geometric configuration; V3 (right) — minor junction.

Primary Study Junction

A junction on the busiest corridor (Figure 2 and Figure 3) in Sonipat, Haryana, India (NCR region) was selected based on: (i) diverse land-use patterns encompassing commercial, residential, and mixed-use zones; (ii) complex traffic composition with mixed motorized and non-motorized vehicles; and (iii) availability of three-year historical crash records (2022–2024).

Validation Junctions

To test model transferability, three additional junctions were selected (Figure 4):

- V1: Different density, different land use
- V2: Similar urban context, different configuration
- V3: Minor junction

Crash Data

Police-reported FIRs for crashes at the junctions (2022–2024) were obtained. A total of 67 crashes were reported at the primary junction.

3.2 Perception Survey & Infrastructure Mapping (Phase-2)

The perception survey was conducted on a voluntary basis. All participants were informed about the study objectives and provided informed consent prior to participation. No personal identifiers were collected, and responses were recorded anonymously. As the study involved non-invasive surveys without personal or sensitive data, formal ethical approval was not required.

3.2.1 Survey and Sample Size Estimation

For this study, power analysis was used to determine the required sample size, as population parameters were unknown. This method ensures

adequate statistical power for detecting meaningful effects (Kemal, 2020) by incorporating effect size, significance level, and desired power. A small effect size of 0.30 was selected to capture the subtle but measurable influence of psychological factors on crash outcomes (Ialongo, 2016), requiring a larger sample for reliability (Simmachan & Boonkrong, 2025). A stringent significance level of 0.01 minimized Type I errors, which are essential considerations in safety research (Erdfelder et al., 1996), while a statistical power of 0.95 ensured a high probability of detecting true effects.

The study evaluated 30 psychological parameters using an F-test ANOVA framework suited for analyzing relationships between multiple predictors and crash dynamics (Erdogan, 2009). Using the chosen parameters, effect size 0.30, significance level 0.01, and power 0.95, the estimated required sample size was 147 respondents. This sample size ensures statistical rigor and supports the study's objectives. With 147 respondents distributed across ~10,000 m² of the primary junction, each respondent represents approximately 68 m², providing sufficient spatial density for generating a reliable KDE surface.

3.2.2 Participant Inclusion Criteria

To ensure reliable perception data, participants met a set of criteria: they had to use the junctions at least three times per week for over six months, be 18–65 years old, and live or work within 2 km of the site. The sample reflected a realistic modal mix of car drivers, two-wheeler riders, pedestrians, and public transport users, ensuring that only regular users with consistent, experience-based perceptions contributed to the study.

3.2.3 Perception Survey Design

The survey comprised two sections: Part A, capturing demographics, and Part B, assessing perceived crash risk.

A key feature was the spatial marking method, in which respondents identified high-risk locations using a digital version of the map. Participants marked points of perceived highest risk (e.g., unsafe crossings, rash driving, poor lighting). Marked points were converted into GPS coordinates, and Kernel Density Estimation (KDE) was applied to generate continuous risk surfaces at 1 m × 1 m resolution. A bandwidth radius of 10 meters was selected, ensuring an optimal balance between spatial smoothness and local sensitivity - large enough to smooth out noise in individual points, yet small enough to preserve fine-grained local variations in perceived risk.

3.2.4 Infrastructure Data Collection

Infrastructure data were collected independently of the perception surveys using systematic GIS mapping to avoid bias. Features included road geometry, lane width, markings, traffic signals, artificial lighting availability, visibility, potholes, and land use. Additional datasets covered traffic volume and speed, building density, points of interest, parking zones, and temporal factors such as rush hours and festivals. This ensured an objective linkage between infrastructure attributes and perceived risk patterns.

3.2.5 Survey Administration and Data Variables

The designed survey was conducted to assess factors influencing crash risk and was administered to 147 participants selected via community centers, markets, and roadsides near the study area. Data collection encompassed a broad range of variables relevant to road safety, including road characteristics (lanes, potholes), contextual factors (land use, point of interest (POIs), population density), and traffic metrics (average daily traffic; ADT, speed, volume-to-capacity ratio) (Ma et al., 2020). Environmental and infrastructural conditions, such as weather, lighting, and visibility, were recorded, along with junction-specific features such as conflict points and pedestrian crossings. The built environment was characterized by building density, street elements, and zone areas (Rankavat & Tiwari, 2015), while additional parameters, including public transport density, red lights, speed limits, and street elements, provided a comprehensive, multi-layered understanding of local traffic and safety conditions.

3.3 Parameters Identification (Phase-3)

The PROMETHEE analysis was applied to identify the most critical parameters influencing perceived crash risk. Attributes such as frequency, type, characteristics, and data availability were prioritized to ensure a robust and practical framework (Goswami, 2020). Frequency captured parameter reliability based on prevalence in literature, while type ensured balanced inclusion of statistical and spatial parameters. Characteristics grouped parameters into physical, psychological, and personal domains, and data availability assessed feasibility on a 1–5 scale. Using PROMETHEE II, a multi-criteria decision analysis method, parameters were systematically ranked through pairwise comparisons and quartile-based frequency distribution (eq. 1–6 in the supplementary material), following normalization.

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \quad eq. 1$$

where, μ_j and σ_j represent the mean and standard deviation of criterion j . A binary preference function was defined as:

$$P(a, b) = \begin{cases} 1 & \text{if } a > b \\ 0 & \text{otherwise} \end{cases} \quad eq. 2$$

to evaluate pairwise preferences between parameters for each criterion. The leaving flow ϕ_i^+ and entering flow ϕ_i^- were calculated as:

$$\phi_i^+ = \frac{1}{n} \sum_{j=1}^n \left(\frac{1}{m} \sum_{k=1}^m P_{ijk} \right) \quad eq. 3$$

$$\phi_i^- = \frac{1}{n} \sum_{j=1}^n \left(\frac{1}{m} \sum_{k=1}^m P_{jik} \right) \quad eq. 4$$

where n is the number of parameters and m is the number of criteria. The net flow ϕ_i , representing overall preference of model i was computed as:

$$\phi_i = \phi_i^+ - \phi_i^- \quad eq. 5$$

The final scores S_i for ranking were obtained by averaging net flow across all criteria, further the parameters were shortlisted:

$$S_i = \frac{1}{m} \sum_{k=1}^m \phi_i \quad eq. 6$$

From 60+ identified variables, 30 parameters were shortlisted. These included behavioral tendencies such as overtaking (with/without medians), tailgating, honking, aggressive driving near exits, and rash driving; infrastructural and environmental factors such as lighting, visibility, congestion, road markings, land use, and seasonal variation; and situational influences such as stress during parking searches, police enforcement, and roadwork zones.

Psychological and context-specific factors, including distraction from billboards or shops, rush-hour stress, perceived psychological risk, pedestrian decision-making, festival influence, and noise levels, were also incorporated along with perceived crash risk. Interactions with vulnerable road users and external elements such as street animals, children crossing, and on-road parking were included to ensure comprehensive crash-risk modeling. Collectively, these parameters strengthen the model by capturing diverse and interconnected contributors to perceived and actual road risk.

3.4 Spatial Analysis (Phase-4)

Infrastructure features were systematically mapped in GIS for the study area at a $1 \text{ m} \times 1 \text{ m}$ grid resolution, including continuous variables such as traffic volume, artificial lighting, and building density; categorical variables such as land use type, presence of crossings, and parking; proximity measures such as distance to the junction and POI density; and derived metrics including volume-capacity ratio and visibility. On-site survey data capturing psychological perceptions of crash risk using a Likert scale (Mukherjee & Mitra, 2020) were spatially linked to these GIS-mapped features. Participants used a map-based questionnaire to mark 30 locations they perceived as high risk for specific factors and questions, rating each on a Likert scale.

These points were converted into Kernel Density Estimation (KDE) surfaces and overlaid onto the centroids of the $1 \text{ m} \times 1 \text{ m}$ grid, assigning perceived risk values to corresponding cells. The regression framework was designed to link survey responses with GIS-derived infrastructure variables. For example, Survey_VC represents the perceived risk associated with the Volume/Capacity Ratio, while Map_VC denotes the corresponding GIS-based measurement derived through kernel density estimation (KDE). Similarly, Map_Ped_Cross and Map_Lighting represent the spatial distribution of pedestrian crossings and street lighting, respectively, extracted from GIS and observational datasets.

The derived regression equations (e.g., $\text{Survey_VC} = 0.07538 + 0.75992 \times \text{Map_VC}$) quantify these relationships, functioning as spatial transfer functions rather than predictive models. They align and scale survey-based perception data with mapped infrastructure on a $1 \text{ m} \times 1 \text{ m}$ grid, enabling consistent spatial comparison. Once standardized, the integrated dataset linking perceived risk and physical infrastructure features was used as the input for the machine learning modeling phase to assess and map crash risk patterns. These equations are mentioned in Table 9 in the supplementary material.

3.5 Modeling (Phase – 5)

Modeling psychological risk in urban road safety requires balancing robustness, interpretability, and the capacity to capture complex non-linear relationships among behavioral, infrastructural, and environmental variables. Artificial Neural Networks (ANN) (eq. 7) model intricate behavioral patterns and spatial heterogeneity but act as black-box systems

with limited interpretability. Support Vector Machines (SVM) (eq. 8) resist overfitting and handle non-linearities effectively through kernel functions, though at the cost of reduced transparency. Decision Trees (DT) (eq. 9) provide clear variable importance rankings and explicit decision logic yet are prone to overfitting without pruning constraints. Integrating these models with Geographic Information Systems (GIS) enhances spatial precision and clarifies how infrastructure and environmental conditions influence perceived crash risk at a $1 \text{ m} \times 1 \text{ m}$ resolution.

An ensemble combining ANN, SVM, and DT was developed to predict spatially distributed perceived crash risk. The complementary strengths of the three architectures, connectionist pattern learning (ANN), margin-based generalization (SVM), and rule-based interpretability (DT), are combined to produce a more robust and transferable prediction system than any single model. Data were normalized and split into an 80:20 train–test ratio, and each model was independently optimized using Randomized Search CV over 5,000 iterations to obtain the best hyperparameter configuration. Final predictions were produced by averaging the outputs from all three optimized models with equal weighting (eq. 10). Equal weighting was adopted to avoid bias toward any single learning paradigm in the absence of prior evidence favoring a specific weighting scheme (Chen & Zheng, 2025). This GIS-integrated ensemble ensures balanced predictive accuracy, stability, and interpretability, offering a robust and transferable framework for modeling urban psychological risk.

The ANN predicts the output $\hat{y}_{ann}$ based on the input features X , using learned weights W and biases b , passed through activation functions f :

$$\hat{y}_{ann} = f(W_2 \cdot f(W_1 \cdot X + b_1) + b_2) \quad eq.7$$

$$\hat{y}_{hybrid} = \frac{(f(W_2 \cdot f(W_1 \cdot X + b_1) + b_2)) + (\sum_{i=1}^n \alpha_i y_i K(X, X_i) + b) + (\sum_{j=1}^J c_j \cdot 1(X \in R_j))}{3} \quad eq. 10$$

where:

W_1 & W_2 = Weight Matrices for the hidden and output layers

b_1 & b_2 = Bias terms for the hidden and output layers

f = Activation Function (ReLU, Tanh, Logistic).

The SVM predicts the output $\hat{y}_{svm}$ using a kernel function $K(X, X_i)$ to map input X into a higher-dimensional space:

$$\hat{y}_{svm} = \sum_{i=1}^n \alpha_i y_i K(X, X_i) + b \quad eq. 8$$

where:

α_i = Lagrange multipliers

y_i = Actual target values for the training samples

$K(X, X_i)$ = Kernel function (e.g., Linear, RBF, Polynomial)

b = Bias term (Constant)

The Decision Tree predicts the output $\hat{y}_{dt}$ by recursively splitting the input space X into regions R_j based on feature thresholds:

$$\hat{y}_{dt} = \sum_{j=1}^J c_j \cdot 1(X \in R_j) \quad eq. 9$$

where:

J : Number of leaf nodes

c_j : Predicted value for region R_j

$X \in R_j$: Indicator function that evaluates to 1 if X belongs to region R_j , otherwise 0.

To enhance predictive accuracy, a hybrid model was developed by combining the predictions of the three optimized models keeping equal weightage. The hybrid prediction $\hat{y}_{hybrid}$ is formulated as:

3.6 Hyperparameters Optimization

The ANN model was optimized using hyperparameters such as hidden layer sizes (10–500 neurons), activation functions (ReLU, tanh, logistic), solvers (adam, SGD), learning rates (constant, adaptive), and iteration limits (500–1000). SVM utilized parameters such as regularization (0.01–1000), kernel types (linear, RBF, poly), and gamma values (scale, auto). For DT, hyperparameters included maximum depth (5–50), minimum samples to split (2–100), and minimum leaf sizes (1–50).

4. Results

4.1 Model Performance and Hyperparameters Optimization Outcome

To model psychological crash risk, an ensemble machine learning framework integrating Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Decision Trees (DT) was developed, with final risk scores derived through weighted averaging of model outputs. This ensemble combined SVM's resistance to overfitting, DT's

interpretability, and ANN's capacity to capture complex non-linear patterns, yielding a balanced prediction system.

After 5,000 iterations, the ensemble achieved strong performance with an R^2 of 0.92 and a cross-validation R^2 of 0.88, demonstrating high predictive accuracy and generalization. The kappa coefficient of 0.779 further indicated substantial agreement between predicted and observed risk categories. Among the individual models, ANN ($R^2 = 0.68$) captured non-linear behavioral responses and spatial heterogeneity using a deeper architecture (79 neurons, SGD). SVM ($R^2 = 0.81$) with a poly kernel ($C = 10$) enhanced stability and minimized overfitting, while DT ($R^2 = 0.94$) contributed interpretability through fine decision splits ($\text{min_samples_leaf} = 2$).

Accuracy across risk levels ranged from 77% to 89% (Figure 5), with the highest accuracy at lower and extreme risk levels (Low: 89%, High: 82%). These results confirmed the ensemble's reliability in predicting psychological crash risk across varying risk levels.

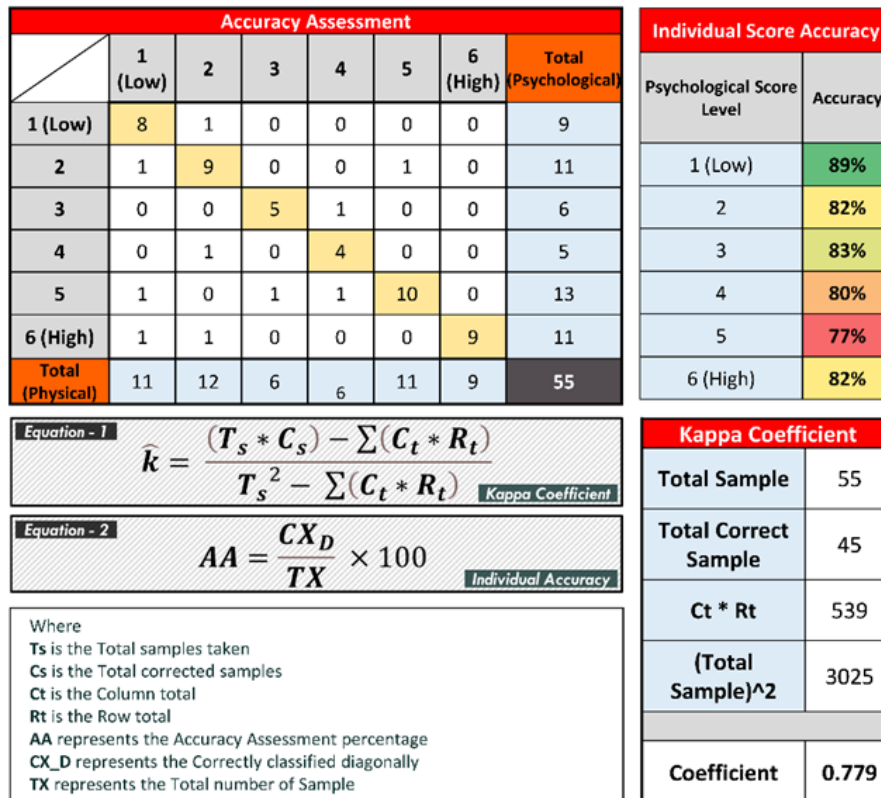


Figure 5 Statistical validation of the ensemble model showing per-class accuracy, kappa coefficient, and individual model performance scores.

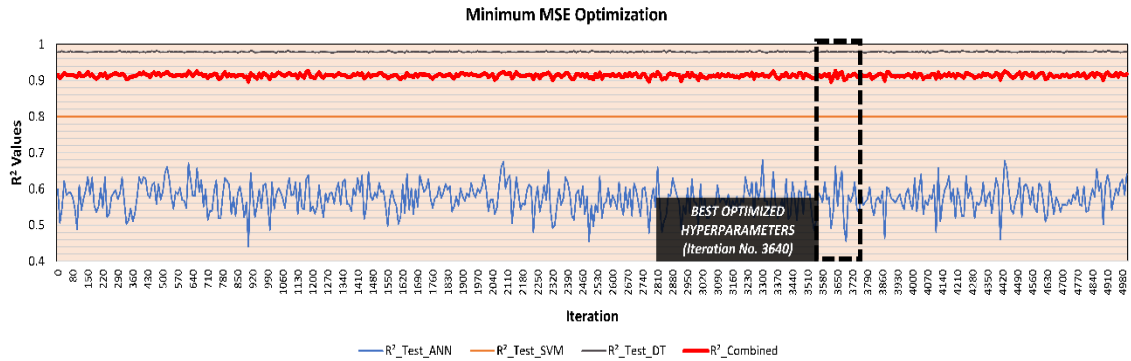


Figure 6 Hyperparameter optimization convergence plot showing R^2 values across 5,000 iterations; optimal configuration reached at iteration 3,640 ($R^2 = 0.92$, MSE minimized).

Figure 6 shows the optimization process over 5,000 iterations, where the best performance occurred at iteration 3640, achieving $R^2 = 0.92$ and cross-validation $R^2 = 0.88$, with the optimized hyperparameter summary in Table 1 and the model comparison in Table 2. While the Decision Tree achieved the highest standalone R^2 , it is less suitable for capturing complex

non-linear relationships and is prone to overfitting (Halabaku & Bytyçi, 2024). The ensemble model (red line) maintained consistently high R^2 values comparable to those of the best-performing individual models, while offering improved robustness, generalization, and transferability by integrating the complementary strengths of ANN, SVM, and DT.

Table 1 Hyperparameter optimization summary

Model	Optimized Hyperparameters	Key Highlights
ANN	{'activation': 'relu', 'solver': 'sgd', 'hidden_layer_sizes': (79), 'learning_rate': 'constant', 'learning_rate_init': 0.001, 'momentum': 0.9, 'max_iter': 908}	Used ReLU activation and SGD solver with moderate hidden layer size (79 neurons) to balance convergence speed and non-linear pattern learning. Momentum (0.9) and constant learning rate (0.001) ensured stability and smooth optimization.
SVM	{'C': 10.0, 'kernel': 'poly', 'gamma': 'scale', 'epsilon': 0.1, 'degree': 3}	The polynomial kernel captured complex relationships, while regularization parameter $C = 10.0$ achieved a balance between bias and variance. The scaled gamma parameter maintained computational stability.
DT	{'criterion': 'squared_error', 'min_samples_split': 5, 'min_samples_leaf': 2, 'splitter': 'best', 'max_depth': None, 'max_features': None, 'max_leaf_nodes': None}	Allowed unrestricted tree depth for maximum flexibility while preventing overfitting using constraints on minimum samples per leaf (2) and per split (5). Used the squared error criterion for accurate regression splitting.

Table 2 Model performance summary

Model	R^2 (Train)	R^2 (Test)	MSE (Test)	Remarks
ANN	0.694	0.662	0.0072	Captures non-linear risk trends; moderate generalization due to adaptive learning.
SVM	0.817	0.799	0.0041	High stability and generalization; robust against overfitting.
DT	0.971	0.942	0.0004	Strong interpretability: slightly high training fit due to flexible depth - risk of overfitting.
Ensemble	0.926	0.880 ^a	0.0109	Best performance at iteration 3640; balanced accuracy, minimal overfitting. ^a Cross-validation R^2 .

^a Cross-validation R^2 reported for the ensemble in place of a single held-out test split

Variables	VIC	Ped Crossing	Art Light	Lanes	Visibility	Wrong Dir	Index Dens	Unauth Park	Child Crossing	Road Driving	LandUse	Impact Intensity	Road Marking	Noise	Advertisement	Slag Density	Obs Stress	Ped Flow	Seasonal Weather	Auth Parking	Police Enforce	Merging	TailGating	Hornring	Petrols	Fair	Street Artery	No Median
VIC	-	0.4002	0.4857	0.7292	0.0447	-0.0667	0.1179	-0.1919	-0.1113	0.4639	-0.5137	0.1667	0.2884	0.0394	0.1319	-0.3307	0.0337	-0.2239	-0.0391	-0.1201	0.3396	0.6606	0.9347	0.0997	-0.1271	0.2155	-0.1569	-0.4843
Ped Crossing	0.4002	-	0.1193	0.4916	0.2031	-0.0074	0.2507	0.1112	0.1762	0.5125	-0.2525	0.3103	0.2249	0.1693	0.3917	-0.2909	0.3822	-0.0485	0.1428	0.0540	0.4682	0.3032	0.4846	0.1030	-0.1342	0.4156	0.2313	0.0962
Art Light	0.4857	0.1193	-	0.1943	-0.0722	0.0016	-0.0808	-0.0768	-0.1950	0.2005	-0.3747	0.0550	0.3274	-0.1441	-0.1288	-0.3714	-0.0896	-0.2100	-0.0729	-0.1734	0.1670	0.3106	0.4860	0.0993	0.2488	-0.0374	0.2511	-0.3777
Lanes	0.7292	0.4916	0.1943	-	0.4403	-0.1158	0.5113	0.0770	0.1229	0.4091	-0.1634	0.1416	0.3198	0.0058	0.3958	-0.1736	0.2311	-0.1347	0.1417	-0.1407	0.3804	0.4474	0.8318	0.1786	-0.1633	0.4809	-0.0266	0.1265
Visibility	0.0447	0.2031	-0.0722	0.4403	-	0.0437	0.8052	0.0951	0.3795	0.5536	0.0473	0.0293	0.0390	0.1542	0.3316	0.0220	0.3270	0.0578	0.3107	0.0384	0.2851	-0.7091	0.1812	0.1067	-0.1129	0.1360	0.1068	0.5158
Wrong Dir	-0.0667	-0.0074	0.0016	-0.1158	0.0437	-	-0.0072	-0.0772	-0.1950	0.2894	0.1909	0.2860	0.0463	0.2424	0.7973	0.2645	0.0460	0.3973	0.2884	0.5648	0.2321	-0.3488	0.0113	0.2294	0.1121	-0.2292	0.4296	0.074
Index Dens	0.1179	0.2507	-0.0808	0.5113	0.8052	-0.0072	-	0.0479	0.3874	0.3295	0.0096	-0.1036	0.0130	0.0672	0.3467	0.0010	0.1983	0.0088	0.3008	-0.0243	0.1950	-0.6531	0.2406	0.1305	-0.1210	0.1574	0.0348	0.3994
UnAuth Park	-0.1919	0.1112	-0.0768	-0.0770	0.0791	0.7172	0.0479	-	0.5296	0.1861	0.5262	0.6241	0.2007	0.6683	0.2945	0.1679	0.4979	0.3235	0.5895	0.3053	0.0631	-0.1617	-0.0705	0.1561	0.1044	-0.0939	0.5006	0.2595
Child Crossing	-0.1113	0.1762	-0.1950	0.1229	0.2705	0.2956	0.3874	0.5296	-	0.2532	0.5954	0.1940	0.2297	0.2719	0.4939	0.5607	0.1957	0.2247	0.2956	0.0831	0.0540	-0.1450	0.0420	0.1859	-0.0304	-0.0395	0.0237	0.3470
Road Driving	0.4639	0.5125	0.2005	0.4091	0.1516	0.3909	0.1335	0.1861	0.0152	-	-0.3629	0.3757	0.2977	0.3023	0.2365	-0.4428	0.4131	-0.0308	0.1387	0.3659	0.5551	0.3753	0.5095	0.0401	0.0393	0.2312	0.2534	0.0074
LandUse	-0.5137	-0.2525	-0.3747	-0.1634	0.0473	0.2690	0.0096	0.3262	0.5954	-0.3629	-	0.0483	0.0362	0.7526	0.3110	0.9497	0.0902	0.2709	0.3481	0.0233	-0.2878	-0.2841	0.2324	0.0830	0.0418	-0.1765	0.1132	0.2295
Impact Intensity	0.1667	0.3103	0.0394	0.1416	0.3198	0.0058	0.0293	0.0390	0.1542	0.3316	0.0220	-	0.2971	0.1815	0.2195	-0.1383	0.6515	0.2751	0.5134	0.4137	0.3091	0.1531	0.2339	0.2640	0.0575	0.1499	0.5257	0.0498
Road Marking	0.2884	0.2249	0.3274	0.3106	0.0384	0.2424	0.0130	0.2897	0.2297	0.2977	0.0362	0.2971	-	0.2102	0.3865	0.0447	0.2725	-0.0099	0.2177	-0.0388	0.1889	0.1256	0.4753	0.2140	0.1533	0.0991	0.3899	0.0832
Noise	0.0394	0.1319	-0.3307	0.0337	-0.2239	-0.0391	-0.1201	0.3396	0.6606	0.9347	0.0997	-0.1271	0.2155	-0.1569	-0.4843	0.1667	0.2884	0.0394	0.1319	-0.3307	0.0337	0.2239	-0.0391	-0.1201	0.3396	0.6606	0.9347	0.0997
Advertisement	0.1319	0.3917	-0.2909	0.3822	-0.0485	0.1428	0.0540	0.4682	0.3032	0.4846	0.1030	-0.1342	0.4156	0.2313	0.0962	0.1667	0.2884	0.0394	0.1319	-0.3307	0.0337	0.2239	-0.0391	-0.1201	0.3396	0.6606	0.9347	0.0997
Slag Density	-0.3307	-0.2909	-0.2909	-0.2909	-0.2909	-0.2909	-0.2909	-0.2909	-0.2909	-0.2909	-0.2909	-0.2909	-0.2909	-0.2909	-0.2909	-0.2909	-0.2909	-0.2909	-0.2909	-0.2909	-0.2909	-0.2909	-0.2909	-0.2909	-0.2909	-0.2909	-0.2909	-0.2909
Obs Stress	0.0337	0.3822	-0.0485	0.1428	0.0540	0.4682	0.3032	0.4846	0.1030	-0.1342	0.4156	0.2313	0.0962	0.1667	0.2884	0.0394	0.1319	-0.3307	0.0337	0.2239	-0.0391	-0.1201	0.3396	0.6606	0.9347	0.0997	-0.1271	0.2155
Ped Flow	-0.2239	-0.0485	-0.2100	-0.1347	0.0578	0.3107	0.0384	0.2851	-0.7091	0.1812	0.1067	-0.1129	0.1360	0.1068	0.5158	0.0447	0.2031	-0.0722	0.4403	-0.0437	0.8052	0.0951	0.3795	0.5536	0.0473	0.0293	0.0390	0.1542
Seasonal Weather	0.0391	0.1428	-0.0729	0.1417	0.3107	0.0384	0.2851	0.0631	-0.1617	0.2406	0.1305	-0.1210	0.1574	0.0348	0.3994	0.0447	0.2031	-0.0722	0.4403	-0.0437	0.8052	0.0951	0.3795	0.5536	0.0473	0.0293	0.0390	0.1542
Auth Parking	0.1670	0.4860	-0.1734	-0.1407	0.3804	0.2851	0.0631	-0.1617	0.2406	0.1305	-0.1210	0.1574	0.0348	0.3994	0.0447	0.2031	-0.0722	0.4403	-0.0437	0.8052	0.0951	0.3795	0.5536	0.0473	0.0293	0.0390	0.1542	0.0074
Police Enforce	0.3032	0.4846	0.1030	0.4682	0.3032	0.4846	0.1030	0.4682	0.3032	0.4846	0.1030	0.4682	0.3032	0.4846	0.1030	0.4682	0.3032	0.4846	0.1030	0.4682	0.3032	0.4846	0.1030	0.4682	0.3032	0.4846	0.1030	0.4682
Merging	0.6606	0.9347	0.0997	0.6606	0.9347	0.0997	0.6606	0.9347	0.0997	0.6606	0.9347	0.0997	0.6606	0.9347	0.0997	0.6606	0.9347	0.0997	0.6606	0.9347	0.0997	0.6606	0.9347	0.0997	0.6606	0.9347	0.0997	0.6606
TailGating	0.9347	0.0997	-0.1271	0.9347	0.0997	-0.1271	0.9347	0.0997	-0.1271	0.9347	0.0997	-0.1271	0.9347	0.0997	-0.1271	0.9347	0.0997	-0.1271	0.9347	0.0997	-0.1271	0.9347	0.0997	-0.1271	0.9347	0.0997	-0.1271	0.9347
Hornring	0.0997	-0.1271	0.2155	0.0997	-0.1271	0.2155	0.0997	-0.1271	0.2155	0.0997	-0.1271	0.2155	0.0997	-0.1271	0.2155	0.0997	-0.1271	0.2155	0.0997	-0.1271	0.2155	0.0997	-0.1271	0.2155	0.0997	-0.1271	0.2155	0.0997
Petrols	-0.1271	0.2155	-0.1569	-0.1271	0.2155	-0.1569	-0.1271	0.2155	-0.1569	-0.1271	0.2155	-0.1569	-0.1271	0.2155	-0.1569	-0.1271	0.2155	-0.1569	-0.1271	0.2155	-0.1569	-0.1271	0.2155	-0.1569	-0.1271	0.2155	-0.1569	-0.1271
Fair	0.2155	-0.1569	-0.0374	0.2155	-0.1569	-0.0374	0.2155	-0.1569	-0.0374	0.2155	-0.1569	-0.0374	0.2155	-0.1569	-0.0374	0.2155	-0.1569	-0.0374	0.2155	-0.1569	-0.0374	0.2155	-0.1569	-0.0374	0.2155	-0.1569	-0.0374	0.2155
Street Artery	0.1569	-0.0374	0.2511	-0.1569	-0.0374	0.2511	-0.1569	-0.0374	0.2511	-0.1569	-0.0374	0.2511	-0.1569	-0.0374	0.2511	-0.1569	-0.0374	0.2511	-0.1569	-0.0374	0.2511	-0.1569	-0.0374	0.2511	-0.1569	-0.0374	0.2511	-0.1569
No Median	-0.4843	0.0962	-0.3777	0.1265	0.5158	0.0774	0.3994	0.2585	0.3470	0.0074	0.2305	0.0408	0.0852	0.1585	0.3865	0.1887	0.4141	0.2160	0.3006	0.0604	0.1169	-0.3492	0.2861	0.0968	-0.0386	0.2820	0.3228	-

Figure 7 Pearson correlation matrix of infrastructure, behavioral, and environmental variables with perceived crash risk scores.

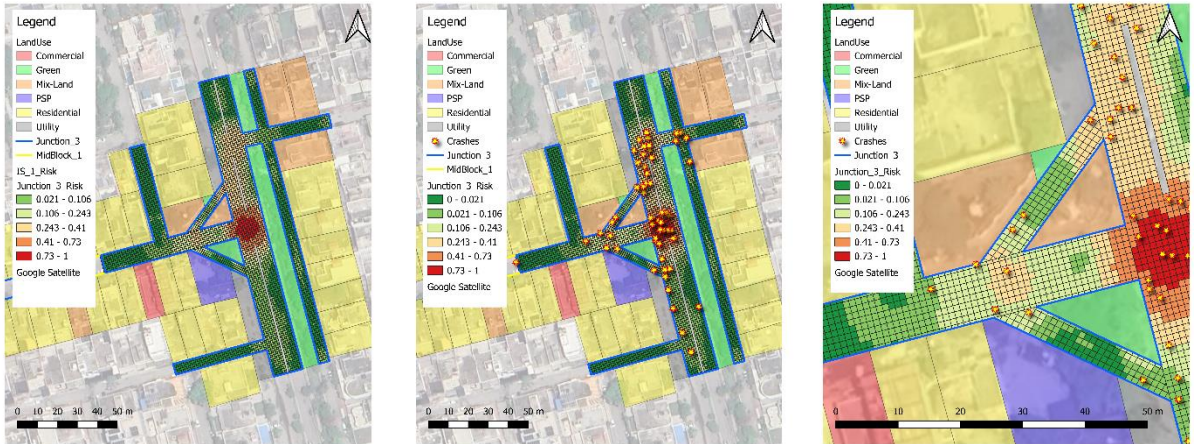


Figure 8 Spatial risk maps for the primary junction: Map 1 (left) — predicted crash risk perception surface; Map 2 (centre) — overlay of historical crash locations on predicted risk; Map 3 (right) — zoomed view of the hidden hazard zone.

4.2 Correlation Analysis and Key Risk Factors

The correlation analysis (Figure 7) shows that perceived crash risk is strongly associated with traffic density, lane configuration, visibility, land-use intensity, and driving behavior. High correlations were observed for tailgating ($r = 0.9347$), building density–land use interaction ($r = 0.9497$), junction density–visibility ($r = 0.8052$), and wrong-direction driving ($r = 0.7371$), while moderate associations were found for pedestrian crossings, unauthorized parking, noise, and seasonal factors. Inverse correlations were identified for artificial lighting, road-marking clarity, and median presence, indicating lower perceived risk where these features were present. Together, these relationships quantitatively characterize how behavioral, infrastructural, and

environmental variables co-vary with psychological crash risk.

4.3 Spatial Risk Distribution at Primary Junction

The developed model was trained and executed for the primary junction, encompassing 4,991 micro-spatial cells at $1\text{ m} \times 1\text{ m}$ resolution. The perceived crash risk, derived from participant responses, was modeled and spatially visualized in Figure 8, Map 1 (left), highlighting variations in perceived safety across the junction area. Map 2 (centre) presents an overlay of actual crash occurrences against perceived risk surfaces. The spatial overlay of perceived and recorded crash risk revealed two distinct clusters. The primary cluster, located near the main junction, showed both high perceived and recorded crash risk,

indicating validated danger zones where users' perceptions accurately reflect actual crash concentrations. In contrast, the secondary cluster exhibited high crash frequency but relatively low perceived risk, reflecting a hidden hazard zone where drivers and pedestrians underestimate danger, as shown in Figure 8, Map 3 (right). Colored zones represent the varying land uses, as shown in the legend of these maps.

Spatial overlay of predicted risk surfaces against recorded crash data revealed four distinct zone types: validated risk zones (high perceived risk AND high crash frequency) covering 16.9% of cells; near-miss zones (high perceived risk, low crash frequency) covering 12.7%; hidden hazard zones (low perceived risk, high crash frequency) covering 6.1%; and safe zones (low perceived risk, low crash frequency) comprising the remaining 64.3%. These proportions confirm that the model captures a spatially coherent pattern rather than random noise, and that a meaningful fraction of the junction area carries elevated risk that is either correctly identified or systematically misperceived by road users.

Validated risk zones (16.9%) represent areas where user perception accurately reflects actual crash concentrations - typically the main junction core, congested lane-merge points, and corridors with poor visibility and high pedestrian activity. These zones confirm the reliability of perception-based risk assessment in identifying genuinely dangerous locations, and represent the highest priority for immediate engineering and enforcement countermeasures. Hidden hazard zones (6.1%), by

contrast, are locations where crash frequency is objectively high yet perceived risk is low. This divergence likely arises from habituation among regular commuters, the absence of visible warning cues, or adaptive behavioral responses that mask underlying danger. These zones are particularly consequential from a planning perspective because conventional awareness campaigns - which target locations users already fear - will not reach them; instead, targeted road-user education and localized investigation of contributing infrastructure factors are required.

Near-miss zones (12.7%) exhibit the inverse pattern: perceived risk is high despite a relatively low recorded crash frequency. These areas reflect the subjective dimension of urban road safety, where poor infrastructure design, environmental discomfort such as noise and crowding, or prior negative experiences elevate perceived danger beyond what crash records alone would suggest. Far from being false alarms, near-miss zones represent proactive intervention opportunities - locations where road users are already signaling discomfort before crashes occur, and where preventive design measures such as traffic calming, improved signage, and pedestrian refuge islands can reduce both fear and future crash likelihood. Collectively, the three non-safe zone types demonstrate that perception-based spatial diagnostics capture safety conditions that crash-record analysis alone cannot reveal, and that each zone type demands a qualitatively different type of intervention. The classification and recommended actions are summarized in Table 3.

Table 3 Crash-Risk Interpretation Summary

Zone Type	% of Cells	Priority	Recommended Actions
Low Crash – Low Risk (Safe Zones)	64.3%	Low	Maintain existing conditions; routine monitoring sufficient.
Low Crash – High Risk (Near-Miss Zones)	12.7%	Medium	Preventive design interventions: traffic calming, improved signage, pedestrian refuge islands, and targeted awareness campaigns.
High Crash – Low Risk (Hidden Hazards)	6.1%	High	Public awareness and road-user education campaigns to correct underestimation of risk; localized infrastructure investigation and enforcement intensification.
High Crash – High Risk (Validated Risk Zones)	16.9%	Urgent	Immediate engineering countermeasures: signal phasing review, junction geometric redesign, median installation, enhanced street lighting, and strict enforcement of traffic violations.

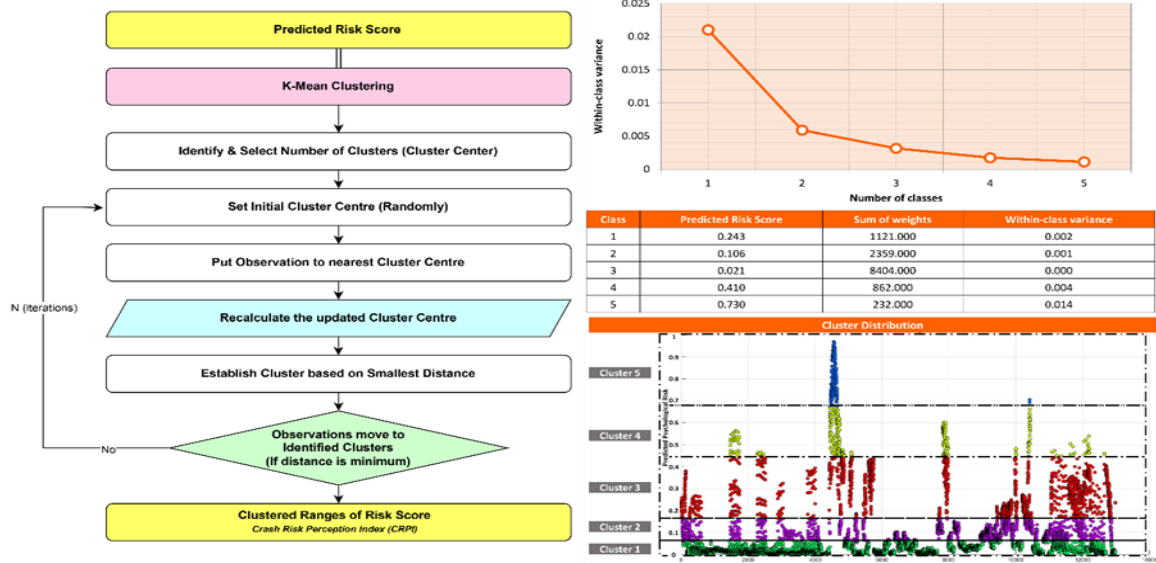


Figure 9 K-means clustering of predicted psychological risk scores into six classes, forming the basis of the Crash Risk Perception Index (CRPI).

The “hidden hazard” phenomenon—where objective crash risk exceeds subjective perception—emerges as a critical focus, emphasizing the need for awareness-based interventions in underestimated danger zones.

4.4 Development of Crash Risk Perception Index (CRPI)

K-means clustering was applied to the psychological risk score data to define clusters of similar risk scores, forming the basis for a proposed "Crash Risk Perception Index," as shown in Figure 9. These clusters enable the classification of safety levels, guiding the implementation of targeted safety

measures to mitigate potential risks. By identifying patterns and trends in the data, this approach provides a structured methodology for grouping data points based on similarity, offering valuable insights for informed decision-making in road safety planning. The choice of six clusters for the 'Crash Risk Perception Index' in Figure 9 is empirically justified based on the k-means clustering algorithm.

The Crash Risk Perception Index (CRPI) was developed to categorize areas based on predicted psychological risk perception scores derived from the ensemble model. The CRPI effectively stratifies risk into six hierarchical safety levels, ranging from A (very safe) to F (critical), as shown in Table 4.

Table 4 Crash Risk Perception Index with Interpretation

CRPI Class	Label	Range (Risk_Score)	Interpretation
1	A (Lowest)	0.000 – 0.021	Represents very safe zones. Perceived and predicted risks are minimal, reflecting stable conditions and strong safety perception.
2	B	0.021 – 0.106	A low-risk class characterized by controlled junctions and moderate pedestrian exposure, with minor perceived discomfort due to occasional congestion.
3	C	0.106 – 0.243	A moderate-risk class with balanced traffic flow. Risk perception remains low but increases in zones with variable land use and minor conflicts.
4	D	0.243 – 0.410	An elevated-risk class where perceived and predicted risks converge. It typically includes areas with poor lane discipline, dense junctions, or mixed traffic composition.
5	E	0.410 – 0.730	A high-risk class strongly associated with high-stress environments, congestion, and recurrent near-miss events. It requires immediate safety interventions.
6	F (Highest)	>0.730	Critical-risk zones representing severe psychological stress and infrastructural vulnerability, often accident-prone spots needing urgent corrective action.

Lower CRPI classes (A–C) correspond to areas with stable traffic behavior, effective infrastructure, and a low perception of danger, whereas higher classes (D–F) reflect stress-inducing traffic environments, poor visibility, and recurrent behavioral violations. Classes E (0.410–0.730) and F (> 0.730) warrant priority interventions, as they represent the upper 15–20% of the perceived and predicted risk spectrum, aligning with spatially validated crash hotspots.

The spatial distribution of the Crash Risk Perception Index (CRPI) across the study area exhibits clear variations in perceived crash risk intensity. A substantial portion of the area lies within the lower-risk categories, with Risk A (0–0.021) comprising 42.66% and Risk B (0.021–0.106) comprising 23.44% of all cells. Moderate-risk zones, represented by Risk C (0.106–0.243) and Risk D (0.243–0.410), account for 14.67% and 14.26% of the area, respectively, while the high-risk categories, E (0.410–0.730) and F (0.730–1), cover only 3.20% and 1.77% of the total area. Although the higher-risk categories (C–F) collectively occupy a smaller spatial extent, these zones correspond to locations with disproportionately greater perceived crash likelihood and safety concerns. Therefore, targeted interventions will prioritize Category F areas as the highest-urgency zones, followed by E, D, and C, ensuring that even a small proportion of the urban space associated with elevated risk receives focused attention. This approach recognizes that a smaller share of the area accounts for a larger share of potential crashes and perceived danger, emphasizing the need for strategic, data-driven interventions to enhance overall junction safety.

4.5 Model Validation Across Multiple Junctions

For model validation, the developed framework was applied to three additional junctions with varying geometric and traffic characteristics to assess its robustness and transferability, as shown in Figure 10. The predicted perceived risk maps were validated through visual overlays of actual crash locations and statistical comparison using the kappa coefficient, which ranged between 0.72 and 0.83, as shown in Table 5, indicating substantial agreement between predicted and observed risk. These results validate the proposed transferable ensemble machine learning framework, confirming that the model demonstrates strong transferability to other junctions with differing characteristics. This strong consistency confirms that the model effectively captures the spatial relationship between perceived and actual risk, demonstrating its applicability to other urban junctions, including those lacking detailed crash records. Across all analyzed sites, including the primary and three validation junctions, the Pearson correlation coefficient between actual crash frequencies and predicted risk values was 0.797. This positive correlation indicates that areas identified by the model as high risk generally correspond to locations with higher recorded crash occurrences. The result reinforces the model's ability to capture realistic spatial relationships between perceived safety and objective crash data, further validating its predictive reliability.

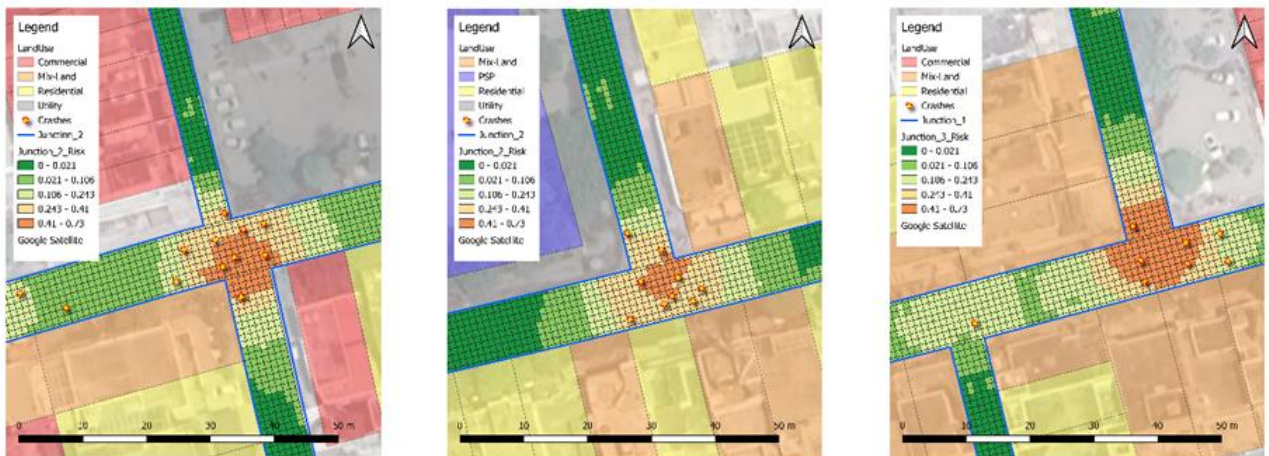


Figure 10 Predicted crash risk surfaces for the three validation junctions: V1 (left), V2 (centre), and V3 (right), used to assess model transferability.

Table 5 Validation junction statistics summary

Junction	Kappa	Total Samples	Overall Accuracy
Main (Primary)	0.779	55	81.82%
V1	0.839	76	86.84%
V2	0.807	64	84.38%
V3	0.723	44	77.27%

Table 6 Perceived risk vs. actual crash relationship statistics summary

Variable	Obs.	Min.	Max.	Mean	Std.Dev.	Correlation (r)	p-value
Psychological Risk (Perceived)	12978	0.000	0.969	0.285	0.183	0.797	<0.001
Crash Density (Actual Observed)	12978	0.000	0.950	0.421	0.276	-	-

Model accuracy was assessed using 239 randomly selected points to compute the kappa coefficient and accuracy for six psychological risk levels. The kappa range of 0.723–0.839 indicated substantial agreement between predicted and observed psychological risk values. Individual accuracy scores ranged from 71% to 92%. These results confirmed the model’s reliability and effectiveness in predicting psychological risk across varying levels.

4.6 Relationship Between Perceived and Actual Crash Risk

To evaluate the alignment between perceived psychological risk and observed crash density, a quantitative correlation analysis was performed using the Pearson correlation coefficient across 12,978 grid cells across all sites, as shown in Table 6. The results demonstrate a strong positive linear association between the two variables, with $r = 0.797$ and $R^2 \approx 0.63$, indicating that perceived risk explains approximately 63% of the spatial variance in actual crash occurrences. The Pearson correlation ($p < 0.001$) further confirms a statistically significant monotonic relationship, reinforcing the robustness of the perceptual measure.

Figure 11 illustrates the statistical validation of perceived crash risk against actual crash frequencies. The regression plot demonstrates a clear upward trend, with higher CRPI values consistently associated with greater crash density. The fitted line closely follows the data points, further reinforcing the strong linear association ($r = 0.797$, $p < 0.001$). This alignment highlights the predictive validity of the CRPI framework and its robustness across spatially diverse urban junctions.

The strong correlation ($r = 0.797$, $R^2 \approx 0.63$) indicates that collective human perception is a robust proxy for actual crash risk, explaining most of the spatial variability in crash density despite underreporting and randomness in crash records. Perceptual judgments compensate for data gaps by reflecting lived exposure to unsafe conditions. The remaining 37% of unexplained variance likely arises from contextual and stochastic factors such as enforcement levels, weather, infrastructure degradation, and near-miss conditions that elevate perceived danger without proportional crashes. Overall, the high correlation affirms the reliability and transferability of perception-based models, particularly in data-scarce or rapidly growing urban settings.

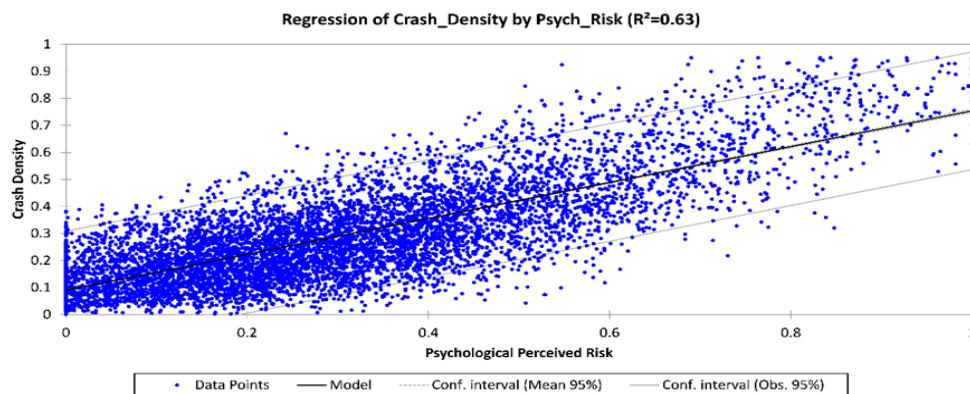


Figure 11 Regression plot of actual crash density against predicted perceived risk (CRPI scores) across all 12,978 grid cells ($r = 0.797$, $p < 0.001$).

These findings show how perception uncovers latent safety conditions not captured by crash statistics alone. The ensemble model's strong alignment between perceptual inputs and crash patterns demonstrates high utility in contexts where crash data are incomplete. Combining psychological perception with machine learning and GIS allows planners to identify emerging hotspots, implement proactive design and traffic management strategies, and support evidence-based safety interventions even in developing or fast-urbanizing regions.

5. Discussion

The findings of this study align with and substantively extend prior research on urban crash risk prediction and perception-informed safety analysis. Consistent with Ma et al. (2020) and Mukherjee & Mitra (2020), perceived risk was found to closely reflect objective crash patterns ($r = 0.797$, $R^2 \approx 0.63$), confirming that collective human interpretation of road conditions constitutes a reliable and spatially coherent proxy for actual safety conditions. This result challenges the prevailing assumption that credible crash prediction depends almost exclusively on formal crash records, and positions human risk perception as an underutilized yet measurable input for evidence-based safety planning - particularly valuable in data-scarce developing-city contexts where police-reported databases are incomplete or inconsistent (Abdulhafedh, 2017).

The ensemble machine learning framework demonstrates clear advantages over single-model approaches. The optimized ANN-SVM-DT ensemble achieved $R^2 = 0.92$ with cross-validation $R^2 = 0.88$ and $\kappa = 0.779$, outperforming each individual model in generalization performance. This is consistent with findings by Berhanu et al. (2023) and Halabaku & Bytyçi (2024), who demonstrate that ensemble methods reduce overfitting and improve transferability by averaging across complementary learning paradigms. The Decision Tree's high standalone test R^2 (0.942) reflects its sensitivity to fine-grained local spatial patterns at $1\text{ m} \times 1\text{ m}$ resolution, while the ensemble's consistently superior κ across all four junctions (0.723–0.839) confirms that robustness - rather than peak training accuracy - is the relevant performance criterion for a transferable safety model.

The spatial clustering analysis yields three distinct zone types with meaningful theoretical and practical implications. Validated danger zones (16.9%

of cells), where both perceived and recorded risk are high, confirm that road users accurately identify the most overtly dangerous conditions - dense junctions with mixed traffic, poor visibility, and high pedestrian activity - consistent with built-environment crash risk findings by Dai (2012) and Rankavat & Tiwari (2016). Hidden hazard zones (6.1% of cells), where crash frequency is high but perceived risk is low, are particularly significant: they represent locations where habituation, familiarity, or the absence of visible warning cues leads road users to underestimate danger, aligning with Andersson's (2011) finding that individuals tend to attribute crash risk to external factors rather than self-assessed vulnerability. Near-miss zones (12.7% of cells), where perceived risk exceeds recorded crash density, reflect the influence of poor infrastructure design, environmental discomfort, and negative past experiences - a pattern consistent with von Stülpnagel & Lucas (2020), who document congruent and incongruent risk sources in urban cycling. Together, these three zone types demonstrate that perception-based diagnostics reveal safety conditions that crash-record analysis alone cannot capture.

The correlation analysis provides a quantitative basis for prioritizing safety interventions. Behavioral factors emerge as the dominant drivers of perceived risk, with tailgating tendency ($r = 0.935$) and building density-land use interaction ($r = 0.950$) representing the strongest associations, underscoring the compounding effect of aggressive driving in spatially dense, mixed-use environments. These findings are consistent with Bucsuházy et al. (2020), who identify driver behavior as the primary human factor in RTC causation, and with Ewing & Dumbaugh (2009), who link mixed-use density to elevated crash likelihood. Conversely, infrastructure improvements show measurable risk-reduction potential: median installation ($r = -0.484$), improved road markings ($r = -0.327$), and enhanced street lighting in dense areas ($r = -0.371$) all significantly reduce perceived danger, consistent with evidence from Retting et al. (2003) on engineering countermeasures and Bivina & Parida (2019) on pedestrian-level safety perception. The full ranking of positive and inverse correlations across behavioral, infrastructural, and environmental categories is summarized in Tables 7 and 8, collectively providing a data-driven foundation for prioritizing junction-level design improvements in cities with limited safety budgets.

Table 7 Category-wise top risk-affecting factors

Category	Factor	r	Interpretation
Behavioral	Tailgating tendency	0.935	Strongest behavioral driver; tailgating drastically heightens anxiety and danger perception.
	Festival noise–stress	0.815	Elevated noise and crowding during festivals amplify psychological stress and perceived risk.
	Wrong driving behavior	0.737	Lane violations and rule-breaking elevate perceived risk, reflecting unsafe driving culture.
	Rash driving	0.737	Aggressive, high-speed driving significantly increases perceived crash risk likelihood.
	Children observed	0.706	Presence of children near roads prompts higher caution and perceived hazard.
Infrastructural	Building density–land use	0.950	Strongest spatial correlation: dense mixed-use development substantially elevates perceived crash risk.
	Visibility–junction density	0.805	Poor visibility at complex junctions intensifies perceived risk.
	Vehicle composition (V/C)	0.661	Mixed vehicle flow and high volume-to-capacity ratios create unstable dynamics and greater perceived danger.
	Unauthorized parking	0.587	Blocked sightlines from unauthorized parking increase perceived risk.
	Pedestrian crossing inadequacy	0.513	Lack of safe crossings heightens perceived threat for pedestrians and drivers alike.
Environmental	Noise–stress interaction	0.732	High ambient noise elevates stress, indirectly increasing perceived crash risk.
	Weather-induced visibility	0.530	Poor weather visibility or waterlogging raises psychological alertness and perceived risk.
	Seasonal advertisements	0.396	Visual clutter from seasonal signage marginally heightens perceived crash potential.
	Pedestrian seasonal flow	0.259	Higher pedestrian activity during certain periods increases perceived danger.

Note: r = Pearson correlation coefficient between the factor and perceived crash risk score

Table 8 Top risk-affecting inversely correlated (protective) factors

Category	Factor	r	Interpretation
Infrastructural	Absence of median – vehicle composition	−0.484	Installing medians decreases perceived collision likelihood and improves traffic flow confidence.
	Road marking clarity	−0.327	Well-maintained lane markings inversely relate to perceived crash risk by enhancing spatial guidance.
Environmental	Artificial lighting – building density	−0.371	Better lighting in dense areas reduces perceived risk, highlighting the role of illumination in safety perception.
	Stray animals – vehicle composition	−0.157	Effective management of stray animals marginally improves psychological comfort in traffic-dense zones.

Note: Negative r values indicate that improvements in these factors are associated with reduced perceived crash risk

Factors that exert an inverse, protective effect on perceived risk - including median installation, road marking clarity, artificial lighting, and stray animal management - are summarized in Table 8. The negative correlations confirm that well-designed infrastructure mitigates perceived danger even in high-density environments, reinforcing the value of targeted physical countermeasures alongside behavioral enforcement strategies.

Some interpretations warrant appropriate caution. The 37% of unexplained spatial variance in crash density likely reflects stochastic and contextual factors - including enforcement levels, transient weather conditions, and temporary traffic disruptions - not captured by the static infrastructure and perception inputs used here. The relationship between environmental stressors such as festival noise or crowding and elevated perceived risk may also reflect broader affective and psychological responses beyond

strict infrastructural deficiencies, suggesting that emotional and attentional dimensions of road safety merit deeper investigation through longitudinal and experimental designs.

Limitations and Future Research Directions

The study's limitations also guide future research priorities. Perception data were collected primarily from regular commuters, who may not fully represent occasional road users, vulnerable groups such as the elderly or mobility-impaired, or those with differing modal patterns, potentially introducing sampling bias into the perceived risk estimates. The temporal snapshot approach restricts the ability to capture seasonal variations, enforcement changes, or short-term fluctuations in traffic that may influence perception over time. While the infrastructural dataset was comprehensive, it does not account for dynamic elements such as temporary road obstructions, pedestrian surges, or transient disruptions. Furthermore, the framework was developed and validated within a single Indian city, which may limit generalizability to contexts with markedly different traffic compositions, cultural road-use norms, or institutional enforcement capacities.

Building on these observations, several priority research directions emerge. Longitudinal perception tracking would establish the temporal stability of psycho-spatial risk patterns and quantify how infrastructure interventions shift both perceived and actual risk over time. Controlled field experiments could isolate the causal effect of specific design changes - lighting upgrades, median installation, or junction simplification - on perceived risk and conflict frequency simultaneously. Testing the framework across diverse international urban contexts, including cities in Southeast Asia, Sub-Saharan Africa, and Latin America, would evaluate its cross-cultural generalizability. Finally, integrating real-time sensor data, connected vehicle trajectories, and pedestrian mobile data could extend the framework from a static diagnostic tool toward a dynamic, continuously updated urban safety monitoring system. Taken together, this study contributes to a growing shift toward viewing crash risk as a psycho-spatial construct shaped jointly by cognition, infrastructure, and behavior - offering a foundation for more holistic, perception-informed urban safety planning.

6. Conclusion

This study developed a comprehensive framework integrating perceived crash risk

assessment, ensemble machine learning modeling, and spatial GIS analysis to predict urban crash risk, particularly in environments with limited formal crash records. The findings show that collective user perception reliably reflects actual crash patterns, demonstrating that human interpretation of road conditions can serve as a practical and effective proxy for objective safety data.

Behavioral, infrastructural, and environmental factors were identified as key determinants of perceived risk, highlighting the combined influence of driver behavior, road design, traffic density, and land-use patterns on urban safety outcomes. Spatial clustering revealed distinct zones where perception and recorded crashes aligned, as well as areas of hidden hazards and near-miss risk, emphasizing the importance of integrating human judgment into risk analysis.

The perception-driven ensemble model, encapsulated in the Crash Risk Perception Index (CRPI), successfully generated high-resolution micro-spatial risk surfaces, enabling planners to prioritize high-risk and near-miss zones for targeted interventions. Overall, this framework offers a scalable, transferable, and data-efficient tool that advances proactive urban safety planning, bridging the gaps among subjective perception, physical infrastructure, and observed crash data.

7. Abbreviations

Abbreviation	Full Term
RTC	Road Traffic Crash
GIS	Geographic Information Systems
KDE	Kernel Density Estimation
PROMETHEE	Preference Ranking Organization Method for Enrichment Evaluation
ANN	Artificial Neural Network
SVM	Support Vector Machine
DT	Decision Tree
CRPI	Crash Risk Perception Index
POI	Point of Interest
ADT	Average Daily Traffic
V/C	Volume-to-Capacity Ratio
FIR	First Information Report
NCR	National Capital Region
R ²	Coefficient of Determination
MSE	Mean Squared Error

8. CRediT Statement

Rohan Vardhan: Conceptualization, Methodology, Software, Data curation, Formal analysis, Visualization, Investigation, Writing – Original draft preparation.

Sewa Ram: Supervision, Methodology, Validation, Writing – Reviewing and Editing.

9. References

- Abdulhafedh, A. (2017). Road traffic crash data: An overview on sources, problems, and collection methods. *Journal of Transportation Technologies*, 7(2), 206-219.
<http://doi.org/10.4236/jtts.2017.72015>
- Adanu, E. K., Smith, R., Powell, L., & Jones, S. (2017). Multilevel analysis of the role of human factors in regional disparities in crash outcomes. *Accident Analysis & Prevention*, 109, 10-17.
<https://doi.org/10.1016/j.aap.2017.09.022>
- Akanbi, O. G., Charles-Owaba, O. E., & Oluleye, A. E. (2009). Human factors in traffic accidents in Lagos, Nigeria. *Disaster Prevention and Management: An International Journal*, 18(4), 397-409.
<https://doi.org/10.1108/09653560910984456>
- Albert, G., & Bekhor, S. (2019). Modelling risky driving behaviour: The role of latent variables in overtaking decision on two-lane highways. *European Journal of Transport and Infrastructure Research*, 19(3), 196-213.
<https://doi.org/10.18757/ejtr.2019.19.3.4385>
- Andersson, H. (2011). Perception of own death risk: An assessment of road-traffic mortality risk. *Risk Analysis: An International Journal*, 31(7), 1069-1082.
<https://doi.org/10.1111/j.1539-6924.2011.01583.x>
- Berhanu, Y., Alemayehu, E., & Schröder, D. (2023). 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. *Journal of Advanced Transportation*, 2023(1), Article 6643412.
<https://doi.org/10.1155/2023/6643412>
- Bivina, G. R., & Parida, M. (2019). Modelling perceived pedestrian level of service of sidewalks: A structural equation approach. *Transport*, 34(3), 339-350.
<https://doi.org/10.3846/transport.2019.9819>
- Black, W. R. (1991). Highway accidents: A spatial and temporal analysis. *Transportation Research Record*, 1318, 75-82.
- Black, W. R., & Thomas, I. (1998). Accidents on Belgium's motorways: A network autocorrelation analysis. *Journal of Transport Geography*, 6(1), 23-31.
[https://doi.org/10.1016/S0966-6923\(97\)00037-9](https://doi.org/10.1016/S0966-6923(97)00037-9)
- Brüde, U., & Larsson, J. (2000). What roundabout design provides the highest possible safety?. *Nordic Road and Transport Research*, 12(2), 17-21.
- Bucsuházy, K., Matuchová, E., Zůvala, R., Moravcová, P., Kostíková, M., & Mikulec, R. (2020). Human factors contributing to the road traffic accident occurrence. *Transportation Research Procedia*, 45, 555-561.
<https://doi.org/10.1016/j.trpro.2020.03.057>
- Cabrera-Arnau, C., Prieto Curiel, R., & Bishop, S. R. (2020). Uncovering the behaviour of road accidents in urban areas. *Royal Society Open Science*, 7(4), Article 191739.
<https://doi.org/10.1098/rsos.191739>
- Chen, S., & Zheng, W. (2025). RRMSE-enhanced weighted voting regressor for improved ensemble regression. *PloS One*, 20(3), Article e0319515.
<https://doi.org/10.1371/journal.pone.0319515>
- Choudhary, J., Ohri, A., & Kumar, B. (2015). Spatial and statistical analysis of road accidents hot spots using GIS [Conference presentation]. *3rd Conference of Transportation Research Group of India (3rd CTRG)*, West Bengal, India.
- Dai, D. (2012). Identifying clusters and risk factors of injuries in pedestrian-vehicle crashes in a GIS environment. *Journal of Transport Geography*, 24, 206-214.
<https://doi.org/10.1016/j.jtrangeo.2012.02.005>
- de Blaeij, A. T., & van Vuuren, D. J. (2003). Risk perception of traffic participants. *Accident Analysis & Prevention*, 35(2), 167-175.
[https://doi.org/10.1016/S0001-4575\(01\)00100-2](https://doi.org/10.1016/S0001-4575(01)00100-2)
- Elander, J., West, R., & French, D. (1993). Behavioral correlates of individual differences in road-traffic crash risk: An examination of methods and findings. *Psychological Bulletin*, 113(2), Article 279.
<https://doi.org/10.1037/0033-2909.113.2.279>

- Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, 28(1), 1-11. <https://doi.org/10.3758/BF03203630>
- Erdogan, S. (2009). Explorative spatial analysis of traffic accident statistics and road mortality among the provinces of Turkey. *Journal of Safety Research*, 40(5), 341-351. <https://doi.org/10.1016/j.jsr.2009.07.006>
- Ewing, R., & Dumbaugh, E. (2009). The built environment and traffic safety: A review of empirical evidence. *Journal of Planning Literature*, 23(4), 347-367. <https://doi.org/10.1177/0885412209335553>
- Goswami, S. S. (2020). Outranking methods: Promethee i and promethee ii. *Foundations of Management*, 12(1), 93-110. <https://doi.org/10.2478/fman-2020-0008>
- Guerrero, T. E., de Dios Ortuzar, J., & Raveau, S. (2020). Traffic accident risk perception among drivers: A latent variable approach. *Transportation Planning and Technology*, 43(3), 313-324. <https://doi.org/10.1080/03081060.2020.1735763>
- Halabaku, E., & Bytyçi, E. (2024). Overfitting in machine learning: A comparative analysis of decision trees and random forests. *Intelligent Automation & Soft Computing*, 39(6), 987-1006. <https://doi.org/10.32604/iasc.2024.059429>
- Hallett, C., Lambert, A., & Regan, M. A. (2012). Text messaging amongst New Zealand drivers: Prevalence and risk perception. *Transportation Research Part F: Traffic Psychology and Behaviour*, 15(3), 261-271. <https://doi.org/10.1016/j.trf.2011.12.002>
- Ialongo, C. (2016). Understanding the effect size and its measures. *Biochemia Medica*, 26(2), 150-163. <https://doi.org/10.11613/BM.2016.015>
- Iamtrakul, P., & Chayphong, S. (2023). Factors affecting the development of a healthy city in Suburban areas, Thailand. *Journal of Urban Management*, 12(3), 208-220. <https://doi.org/10.1016/j.jum.2023.04.002>
- Jadaan, K., Al-Braizat, E., Al-Rafayah, S., Gammoh, H., & Abukahlil, Y. (2018). Traffic safety in developed and developing countries: A comparative analysis. *Journal of Traffic and Logistics Engineering*, 6(1), Article 157. <https://doi.org/10.18178/jtle.6.1.1-5>
- Kam, B. H. (2003). A disaggregate approach to crash rate analysis. *Accident Analysis & Prevention*, 35(5), 693-709. [https://doi.org/10.1016/S0001-4575\(02\)00048-9](https://doi.org/10.1016/S0001-4575(02)00048-9)
- Keken, Z., Sedonik, J., Kušta, T., Andrášik, R., & Bil, M. (2019). Roadside vegetation influences clustering of ungulate vehicle collisions. *Transportation Research Part D: Transport and Environment*, 73, 381-390. <https://doi.org/10.1016/j.trd.2019.07.013>
- Kemal, Ö. (2020). Power analysis and sample size, when and why?. *Turkish Archives of Otorhinolaryngology*, 58(1), 3-4. <https://doi.org/10.5152/tao.2020.0330>
- Legree, P. J., Heffner, T. S., Psotka, J., Martin, D. E., & Medsker, G. J. (2003). Traffic crash involvement: Experiential driving knowledge and stressful contextual antecedents. *Journal of Applied Psychology*, 88(1), 15-26. <https://doi.org/10.1037/0021-9010.88.1.15>
- Levine, N., Kim, K. E., & Nitz, L. H. (1995). Spatial analysis of Honolulu motor vehicle crashes: I. Spatial patterns. *Accident Analysis & Prevention*, 27(5), 663-674. [https://doi.org/10.1016/0001-4575\(95\)00017-T](https://doi.org/10.1016/0001-4575(95)00017-T)
- Lund, I. O., & Rundmo, T. (2009). Cross-cultural comparisons of traffic safety, risk perception, attitudes and behaviour. *Safety Science*, 47(4), 547-553. <https://doi.org/10.1016/j.ssci.2008.07.008>
- Ma, Y., Meng, H., Chen, S., Zhao, J., Li, S., & Xiang, Q. (2020). Predicting traffic conflicts for expressway diverging areas using vehicle trajectory data. *Journal of Transportation Engineering, Part A: Systems*, 146(3), Article 4020003. <https://doi.org/10.1061/JTEPBS.0000320>
- Machado-León, J. L., de Oña, J., de Oña, R., Eboli, L., & Mazzulla, G. (2016). Socio-economic and driving experience factors affecting drivers' perceptions of traffic crash risk. *Transportation Research Part F: Traffic Psychology and Behaviour*, 37, 41-51. <https://doi.org/10.1016/j.trf.2015.11.010>
- McKenna, F. P., Horswill, M. S., & Alexander, J. L. (2006). Does anticipation training affect drivers' risk taking?. *Journal of Experimental Psychology: Applied*, 12(1), Article 1. <https://psycnet.apa.org/doi/10.1037/1076-898X.12.1.1>

- Mukherjee, D., & Mitra, S. (2020). Modelling risk factors for fatal pedestrian crashes in Kolkata, India. *International Journal of Injury Control and Safety Promotion*, 27(2), 197-214. <https://doi.org/10.1080/17457300.2020.1725894>
- Özkan, T., Lajunen, T., Doğruyol, B., Yıldırım, Z., & Çoymak, A. (2012). Motorcycle accidents, rider behaviour, and psychological models. *Accident Analysis & Prevention*, 49, 124-132. <https://doi.org/10.1016/j.aap.2011.03.009>
- Panimalar, S. A., & Krishnakumar, A. (2023). A review of churn prediction models using different machine learning and deep learning approaches in cloud environment. *Journal of Current Science and Technology*, 13(1), 136-161. <https://doi.org/10.14456/jcst.2023.12>
- Petch, R. O., & Henson, R. R. (2000). Child road safety in the urban environment. *Journal of Transport Geography*, 8(3), 197-211. [https://doi.org/10.1016/S0966-6923\(00\)00006-5](https://doi.org/10.1016/S0966-6923(00)00006-5)
- Petridou, E., & Moustaki, M. (2000). Human factors in the causation of road traffic crashes. *European Journal of Epidemiology*, 16(9), 819-826. <https://doi.org/10.1023/A:1007649804201>
- Prasannakumar, V., Vijith, H., Charutha, R., & Geetha, N. (2011). Spatio-temporal clustering of road accidents: GIS based analysis and assessment. *Procedia-Social and Behavioral Sciences*, 21, 317-325. <https://doi.org/10.1016/j.sbspro.2011.07.020>
- Ram, T., & Chand, K. (2016). Effect of drivers' risk perception and perception of driving tasks on road safety attitude. *Transportation Research Part F: Traffic Psychology and Behaviour*, 42, 162-176. <https://doi.org/10.1016/j.trf.2016.07.012>
- Rankavat, S., & Tiwari, G. (2015). Association between built environment and pedestrian fatal crash risk in Delhi, India. *Transportation Research Record*, 2519(1), 61-66. <https://doi.org/10.3141/2519-07>
- Rankavat, S., & Tiwari, G. (2016). Pedestrians risk perception of traffic crash and built environment features—Delhi, India. *Safety Science*, 87, 1-7. <https://doi.org/10.1016/j.ssci.2016.03.009>
- Retting, R. A., Ferguson, S. A., & McCartt, A. T. (2003). A review of evidence-based traffic engineering measures designed to reduce pedestrian–motor vehicle crashes. *American Journal of Public Health*, 93(9), 1456-1463. <https://doi.org/10.2105/AJPH.93.9.1456>
- Sajed, Y., Shafabakhsh, G., & Bagheri, M. (2019). Hotspot location identification using accident data, traffic and geometric characteristics. *Engineering Journal*, 23(6), 191-207. <https://doi.org/10.4186/ej.2019.23.6.191>
- Shahi, S., Brussel, M., & Grigolon, A. (2023). Spatial analysis of road traffic crashes and user based assessment of road safety: A case study of Rotterdam. *Traffic Injury Prevention*, 24(7), 567-576. <https://doi.org/10.1080/15389588.2023.2234530>
- Simmachan, T., & Boonkrong, P. (2025). Effect of resampling techniques on machine learning models for classifying road accident severity in Thailand. *Journal of Current Science and Technology*, 15(2), Article 99. <https://doi.org/10.59796/jcst.V15N2.2025.99>
- Sjöberg, L., Moen, B. E., & Rundmo, T. (2004). Explaining risk perception. An evaluation of the psychometric paradigm in risk perception research. *Rotunde Publikasjoner Rotunde*, 84, 55-76.
- Soltani, A., & Askari, S. (2017). Exploring spatial autocorrelation of traffic crashes based on severity. *Injury*, 48(3), 637-647. <https://doi.org/10.1016/j.injury.2017.01.032>
- Steenberghen, T., Dufays, T., Thomas, I., & Flahaut, B. (2004). Intra-urban location and clustering of road accidents using GIS: A Belgian example. *International Journal of Geographical Information Science*, 18(2), 169-181. <https://doi.org/10.1080/13658810310001629619>
- Vergel-Tovar, C., López, S., Lleras, N., Hidalgo, D., Rincon, M., Orjuela, S., & Vega, J. (2020). Examining the relationship between road safety outcomes and the built environment in Bogotá, Colombia. *Journal of Road Safety*, 31(3), 33-47. <https://doi.org/10.33492/JRS-D-20-00254>
- Verma, N. K., Chaudhary, M. H. K., & Soni, M. D. K. (2022). A Study of on-Street Parking Vehicle in Sonipat City. *International Journal for Science Technology and Engineering*, 10, 37-42. <https://doi.org/10.22214/ijraset.2022.46948>
- von Stülpnagel, R., & Lucas, J. (2020). Crash risk and subjective risk perception during urban cycling: Evidence for congruent and incongruent sources. *Accident Analysis &*

- Prevention*, 142, Article 105584.
<https://doi.org/10.1016/j.aap.2020.105584>
- Wang, X., Peng, Y., Yi, S., Wang, H., & Yu, W. (2021a). Risky behaviors, psychological failures and kinematics in vehicle-to-powered two-wheeler accidents: Results from in-depth Chinese crash data. *Accident Analysis & Prevention*, 156, Article 106150.
<https://doi.org/10.1016/j.aap.2021.106150>
- Wang, M., Yi, J., Chen, X., Zhang, W., & Qiang, T. (2021b). Spatial and Temporal distribution analysis of traffic accidents using GIS-Based data in Harbin. *Journal of Advanced Transportation*, 2021(1), Article 9207500.
<https://doi.org/10.1155/2021/9207500>
- World Health Organization. (2021). *WHO Kicks Off a Decade of Action for Road Safety*. Retrieved from <https://www.who.int/news/item/28-10-2021-who-kicks-off-a-decade-of-action-for-road-safety>
- Xiao, T., Lu, H., Wang, J., & Wang, K. (2021). Predicting and interpreting spatial accidents through MDLSTM. *International Journal of Environmental Research and Public Health*, 18(4), Article 1430.
<https://doi.org/10.3390/ijerph18041430>
- Xie, Z., & Yan, J. (2008). Kernel density estimation of traffic accidents in a network space. *Computers, Environment and Urban Systems*, 32(5), 396-406.
<https://doi.org/10.1016/j.compenvurbsys.2008.05.001>
- Yao, S., Loo, B. P., & Yang, B. Z. (2016). Traffic collisions in space: Four decades of advancement in applied GIS. *Annals of GIS*, 22(1), 1-14.
<https://doi.org/10.1080/19475683.2015.1085440>
- ZahranEl-Said, M. M., Jiann, T. S., Mohamad'Asri, N. A. A. B., Tan, E. H. A., Yap, Y. H., & Rahman, E. K. A. (2019). Evaluation of various GIS-based methods for the analysis of road traffic accident hotspot. *MATEC Web of Conferences*, 258, Article 3008.
<https://doi.org/10.1051/mateconf/201925803008>
- Zhang, Y., Jing, L., Sun, C., Fang, J., & Feng, Y. (2019). Human factors related to major road traffic accidents in China. *Traffic Injury Prevention*, 20(8), 796-800.
<https://doi.org/10.1080/15389588.2019.1670817>