iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** XI **Month of publication:** November 2026

DOI: https://doi.org/10.22214/ijraset.2026.85625

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue XI Nov 2026- Available at www.ijraset.com

A Data-Driven Assessment and Prioritization of Urban Road Accident Burden in Indian Cities for Transportation Safety

Macha Anantha Lakshmi[1], B. Vinod Kumar[2], T. Kiran Kumar[3]
[1]*M.Tech. Scholar, Department of Civil Engineering, St. Ann's College of Engineering & Technology, Chirala, Andhra Pradesh – 523187, India.*
[2]*Assistant Professor, Department of Civil Engineering, St. Ann's College of Engineering & Technology, Chirala, Andhra Pradesh – 523187, India.*
[3]*Professor, Department of Civil Engineering, St. Ann's College of Engineering & Technology, Chirala, Andhra Pradesh – 523187, India.*

Abstract: ***Urban road-safety planning requires a transparent basis for identifying where engineering, enforcement and public-awareness resources may be prioritized. This study develops a reproducible framework to assess reported road accident burden across 50 Indian million-plus cities using official statistics. The 2023 city-level dataset contains 81,144 reported accidents, 17,255 fatalities and 73,476 injuries. The framework combines aggregate trend analysis, entropy-weighted Technique for Order Preference by Similarity to Ideal Solution (TOPSIS), K-means clustering, ranking sensitivity analysis and a two-axis burden–severity matrix. Entropy weighting assigns weights of 0.3412, 0.2695 and 0.3894 to accidents, fatalities and injuries, respectively. Delhi, Bengaluru and Jabalpur are the three highest-ranked cities under the composite burden score. A five-cluster solution yields a silhouette coefficient of 0.4387 and mean pairwise adjusted Rand index of 0.9929 across 20 initializations. Rank agreement is high under equal, accident-focused and injury-focused weights, but lower when fatalities receive dominant emphasis. The results show that city prioritization depends on the selected outcome dimension. The framework is a screening-level decision-support approach: because consistent city-specific exposure denominators are unavailable, the results represent reported burden and must not be interpreted as exposure-adjusted crash risk or causal evidence.***

Keywords: Urban road safety; accident burden; Indian cities; entropy weighting; TOPSIS; K-means; sensitivity analysis; official statistics.

I. INTRODUCTION

Road traffic injuries remain a major public health and transport planning challenge. The World Health Organization estimates approximately 1.19 million annual road traffic deaths worldwide, with a disproportionate burden in low- and middle-income countries [1]. The Global Plan for the Decade of Action 2021–2030 calls for coordinated action across road-safety management, safer roads, safer vehicles, safer road users and post-crash response [2]. India recorded 480,583 road accidents, 172,890 fatalities and 462,825 injuries in 2023 [3]. Urban transport systems are especially complex because pedestrians, cyclists, two-wheelers, public transport and private vehicles share constrained road space. A single-variable ranking may conceal important differences: a large city may record more crashes because of its population and travel demand, whereas another may show a higher fatality count relative to reported accidents. Valid risk comparisons require exposure measures such as vehicle-kilometers travelled, traffic volume, population or road length [4,5,15]. This study therefore focuses on reported burden rather than intrinsic risk.

The objectives are to describe the 2019–2023 aggregate trend for the 50-city group; rank cities using accident, fatality, and injury counts; identify descriptive burden profiles; test the sensitivity of rankings to alternative weights; and distinguish absolute accident burden from fatality severity.

II. RELATED WORK AND RESEARCH GAP

Road-safety assessment has evolved from simple crash counts towards integrated systems that consider outcomes, exposure, road-user behaviour, infrastructure and post-crash care. The WHO road-safety data manual emphasizes consistent definitions, complete outcome data and exposure indicators for valid comparisons [4]. Barffour et al. [5] identified limitations in publicly available Indian road-safety data, including uneven availability of exposure and safety-performance measures. These constraints remain important when secondary statistics are used to compare cities.

©IJRASET: All Rights are Reserved

Indian urban safety studies provide detailed evidence on vulnerable road users and local crash patterns. Mohan et al. [6] examined fatal crashes in six Indian cities and highlighted the contribution of pedestrians, cyclists and motorized two-wheeler users. Paul [7] investigated motorcycle crash rates at unsignalized urban intersections, demonstrating the value of site-level exposure and behavioral variables. Such studies provide diagnostic detail but require local data that are not consistently available across a large set of cities.

Multivariate and spatial approaches can identify groups of locations with similar crash characteristics. Anderson [8] combined kernel density estimation and K-means clustering to profile crash hotspots. Maji et al. [9] proposed a clustering framework to identify benchmark and critical regions for regional safety strategies. Multi-criteria decision-making methods provide another route for combining indicators measured in different units. TOPSIS ranks alternatives according to their distance from positive and negative ideal solutions [10], while entropy weighting derives criterion weights from the dispersion of observed data [11]. Since data-derived weights do not necessarily represent policy priorities, sensitivity analysis is required.

The research gap addressed here is the lack of a compact, reproducible city-screening workflow that combines composite burden ranking, cluster profiling, rank sensitivity and a separate severity dimension while explicitly respecting the geographic limits of official datasets. The contribution is methodological transparency, not a new causal crash model.

III. STUDY AREA, DATA AND ANALYTICAL FRAMEWORK

The primary source is the Ministry of Road Transport and Highways (MoRTH), Road Accidents in India 2023 [3]. Table 6.2 provides accident, fatality and injury counts for 50 million-plus cities in 2023. Table 6.1 provides annual aggregate totals for the same city group from 2019 to 2023. These annual values are group totals, not a city-by-year panel. National traffic-violation, road-feature, and junction tables are not joined to city records because their spatial aggregations differ.

The 50-city totals reconcile to 81,144 accidents, 17,255 fatalities and 73,476 injuries. Table 1 defines the datasets and their analytical use.

Table 1. Source datasets, coverage and role in the analysis.

Source table	Coverage	Analytical role
MoRTH Table 6.2	50 cities; 2023 accident, fatality and injury counts	Primary cross-section for ranking, clustering and prioritization
MoRTH Table 6.1	50-city group aggregate; 2019–2023	Temporal context; not a city-level panel
MoRTH Tables 3.1, 3.5 and 3.6	National violation, road-feature and junction totals	Context only; not city-level explanatory variables
MoRTH Table 6.10	Categorized junction accidents in 50 million-plus cities	Separate junction context; does not cover all city accidents

The workflow comprises source extraction and reconciliation, indicator construction, multi-criteria ranking, clustering, sensitivity testing and burden–severity classification. All computations were performed in Python; the analyses are descriptive and exploratory.

IV. MATHEMATICAL FORMULATION

For each city i (i = 1,…,50), let A_i, F_i and I_i denote reported accidents, fatalities and injuries. Fatality severity is represented descriptively as:

$$S_i = 100 \times F_i / A_i$$

S_i is the number of fatalities per 100 reported accidents. It is not the probability of death conditional on an individual crash, nor does it adjust for traffic exposure.

A. Entropy-weighted TOPSIS

The three-count criteria are treated as benefit criteria because higher values indicate a greater recorded burden. For entropy weighting, $p_{ij} = r_{ij} / \Sigma_i r_{ij}$; $e_j = -(1/\ln n)\Sigma_i p_{ij} \ln(p_{ij})$; and $w_j = (1-e_j)/\Sigma_j(1-e_j)$, where n = 50. TOPSIS vector-normalizes each criterion, multiplies it by its weight and calculates Euclidean distances from the positive and negative ideal solutions. The relative closeness coefficient is $C_i = D_i^-/(D_i^+ + D_i^-)$; a higher value indicates a higher composite burden within the sample [10,11].

B. K-means clustering

For clustering, [A_i, F_i, I_i, S_i] are transformed using log(1+x) and standardized to zero mean and unit variance. K-means minimizes the within-cluster sum of squared Euclidean distances, $J = \Sigma_k \Sigma_{i \in C_k} \| x_i - \mu_k \|^2$ [12]. Candidate k values from 2 to 10 are evaluated using the silhouette coefficient [13]. The final k = 5 solution is selected as a parsimonious profile, rather than solely maximizing silhouette, because k = 6 and 7 create very small groups. Stability is assessed using pairwise adjusted Rand indices (ARI) across 20 initializations [14].

C. Weight sensitivity and burden–severity matrix

TOPSIS rankings are recalculated for equal weights and three focused vectors: accident emphasis (0.6, 0.2, 0.2), fatality emphasis (0.2, 0.6, 0.2) and injury emphasis (0.2, 0.2, 0.6). Spearman rank correlation and top-ten overlap are compared with the entropy ranking. A separate two-axis matrix uses the 75th-percentile thresholds of accident counts and S_i to distinguish high versus lower burden and high versus lower severity.

V. EXPERIMENTAL DESIGN AND COMPUTATIONAL SETUP

This is a retrospective secondary data analysis, not a physical experiment or a traffic microsimulation. Data preparation and analysis were performed using Python with pandas, NumPy, scikit-learn, SciPy, and Matplotlib. K-means used 100 initializations for the final fit and a fixed random state of 2026. Candidate-k evaluation used silhouette coefficients; stability was assessed over 20 initializations (190 pairwise comparisons). No crash prediction, causal inference or field validation was performed.

The analysis pipeline preserves separate geographic units: city-level data support city comparisons; national contributing-factor data provide only country-level context. Missing exposure denominators are not imputed.

VI. RESULTS AND DISCUSSION

A. Aggregate trend for the 50-city group

Reported accidents in the 50-city group declined from 82,781 in 2019 to 58,736 in 2020 and then increased to 81,144 in 2023. Fatalities declined from 17,798 in 2019 to 13,542 in 2020, then rose to 17,255 in 2023. The reduction in 2020 coincides with the COVID-19 mobility disruption and should not be interpreted as a structural safety improvement without accounting for changes in travel exposure. By 2023, accident and fatality totals had approached their 2019 levels.

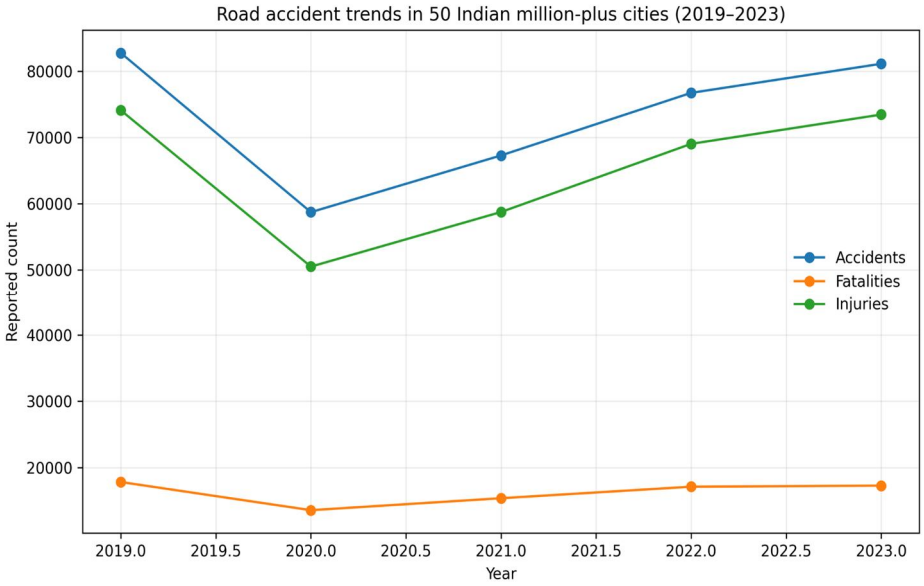


Figure 1. Aggregate road accidents, fatalities and injuries in the 50 million-plus city group, 2019–2023.

Figure 1 presents the group-level trend. It does not show individual city trajectories because the source table reports only annual aggregate totals for the group.

B. Entropy weighting and city ranking

Entropy weights were 0.3412 for accidents, 0.2695 for fatalities and 0.3894 for injuries. The higher injury weight reflects greater dispersion in the observed data; it does not imply greater normative importance. Delhi ranked first (C_i = 1.000), followed by Bengaluru (0.735) and Jabalpur (0.643). Chennai and Malappuram completed the top five. Table 2 reports the ten highest-ranked cities. Figure 2 visualizes the same ordering; the table is retained for exact values and the figure for rapid comparison.

Table 2. Ten highest-priority cities under entropy-weighted TOPSIS.

Rank	City	Score	Rank	City	Score
1	Delhi	1.000	6	Jaipur	0.482
2	Bengaluru	0.735	7	Indore	0.441
3	Jabalpur	0.643	8	Hyderabad	0.405
4	Chennai	0.566	9	Thrissur	0.404
5	Malappuram	0.500	10	Mumbai	0.392

These scores combine absolute counts of accidents, fatalities, and injuries. The ranking therefore indicates the relative scale of reported burden within the selected 50-city sample; it is not a per-capita, per-vehicle or per-kilometer risk ranking.

C. Descriptive city-burden profiles

The five-cluster solution achieved a silhouette coefficient of 0.4387. Mean pairwise ARI across 20 initializations was 0.9929, with a minimum of 0.9625, indicating high stability under the tested initializations. The cluster profiles are descriptive groupings based on the selected variables and preprocessing choices; they are not validated safety grades.

Table 3 summarizes cluster sizes and mean accident counts. Profile 5 contains six cities and has the largest mean accident count (4,019), while Profile 1 contains four cities and has the smallest (349). Profile 2 has 15 cities and a mean accident count of 542. The profile heatmap in Figure 3 shows the relative pattern of the four clustering variables after within-indicator min–max scaling.

Table 3. Summary of the five K-means city-burden profiles.

Profile	Number of cities	Mean reported accidents
1	4	349
2	15	542
3	14	1,375
4	11	2,569
5	6	4,019

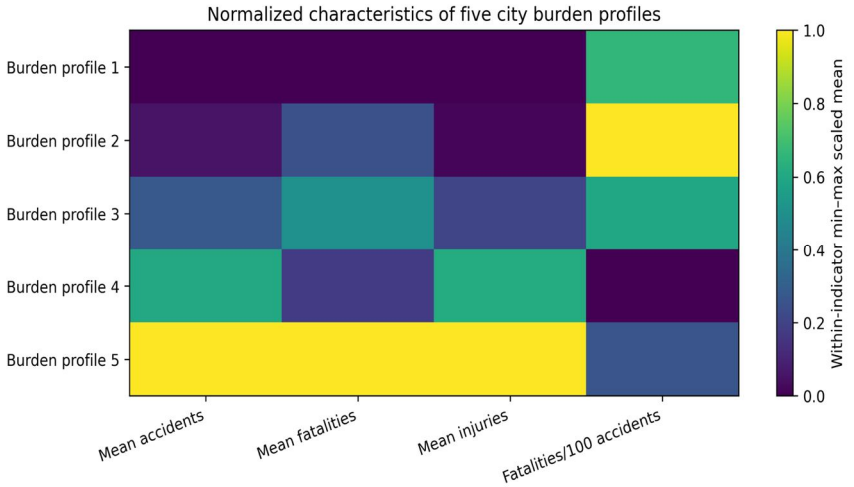


Figure 2. Within-indicator min–max scaled mean characteristics of the five city-burden profiles.

The heatmap is intended to compare relative profile shapes, not absolute magnitudes across indicators. The cluster labels are arbitrary identifiers and should not be interpreted as an ordinal safety scale. A separate silhouette plot is not included in the main manuscript because the numerical selection result and stability statistics are reported directly in the text; it can be retained as supplementary material.

D. Interpretation for planning

The clustering and ranking answer different questions. TOPSIS provides a single composite ordering, whereas clustering groups cities with similar multivariate patterns. Neither method identifies crash locations or the engineering mechanisms behind the observed counts. City agencies should use the profiles to screen for further investigation, then consult local crash records, exposure data, and road-safety audits before selecting interventions.

E. Sensitivity of city ranking to indicator weights

The entropy ranking showed Spearman correlations of 0.993 under equal weights, 0.993 under accident emphasis, 0.979 under injury emphasis and 0.776 under fatality emphasis. Top-ten overlap with the entropy ranking was 10/10, 8/10, 9/10 and 5/10, respectively. This indicates that the ordering is comparatively stable when accidents or injuries receive greater emphasis, but changes more noticeably when fatalities dominate the composite score.

Table 4. Ranking sensitivity relative to entropy-weighted TOPSIS.

Weight scenario	Spearman ρ	Top-10 overlap	Interpretation
Equal weights	0.993	10/10	Very high agreement
Accident emphasis	0.993	8/10	Very high agreement
Injury emphasis	0.979	9/10	Very high agreement
Fatality emphasis	0.776	5/10	Moderate agreement

Figure 4 displays rank changes for selected cities and illustrates why a composite ranking should be accompanied by sensitivity analysis. In particular, cities that perform strongly on accident and injury counts may rank lower when the fatality count is assigned the dominant weight. Weight selection should therefore be reported transparently and aligned with the decision objective.

F. Two-axis prioritization by burden and severity

The 75th-percentile thresholds were 2,208.25 reported accidents and 42.9489 fatalities per 100 accidents. The matrix classified 13 cities as high burden/lower severity, 13 as lower burden/high severity, 24 as lower burden/lower severity and none as high burden/high severity. Figure 5 visualizes this separation.

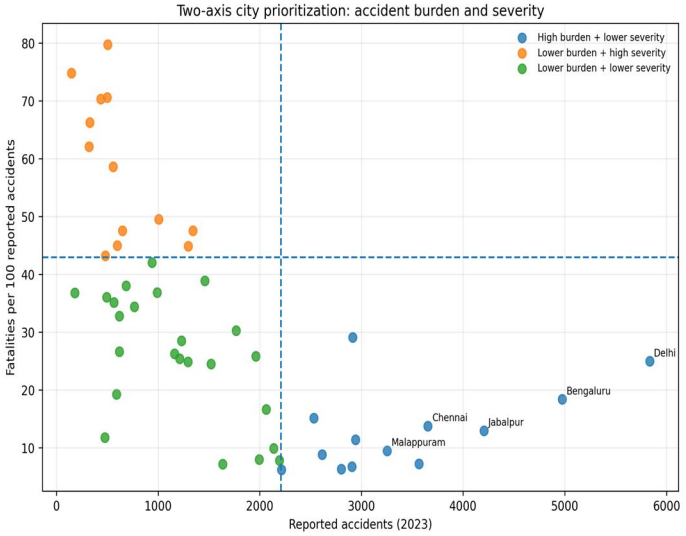


Figure 3. City classification using the 75th-percentile thresholds for 2023 accident count and fatalities per 100 reported accidents.

The two-axis result prevents high-volume cities from automatically being treated as the most severe. Cities in the lower-burden/high-severity group may warrant closer review of crash circumstances and post-crash response, while high-burden/lower-severity cities may require investigation of frequent collision events and exposure. The absence of a high–high quadrant is threshold- and sample-dependent; it does not indicate an absence of serious road-safety problems.

The national violation, road-feature and junction distributions were not used to explain city ranks because they are not city-specific explanatory variables. This separation avoids ecological inference and unsupported causal interpretation.

VII. LIMITATIONS AND FUTURE WORK

The analysis has five principal limitations. First, it uses police-reported aggregate counts, which may be affected by under-reporting and differences in recording practices. Second, the selected city tables do not provide consistent denominators for population, registered vehicles, traffic volume, vehicle-kilometers or road length; exposure-adjusted rates therefore cannot be calculated. Third, the fatality-to-accident ratio is a descriptive aggregate indicator and should not be read as crash-level fatality probability. Fourth, national traffic-violation, road-feature and junction distributions cannot be attributed to individual cities. Fifth, clustering depends on the selected indicators, transformation, scaling and cluster count; external validation has not been undertaken.

Future work should integrate city-specific exposure measures and geocoded crash records. Population, vehicle registrations, traffic counts, road-network length and vehicle-kilometers would support more defensible comparative rates. Geocoded records would enable road-segment and intersection hotspot analysis, spatial modelling and identification of vulnerable road users. Linking police records with hospital and emergency-response datasets could improve the completeness of injury-severity data. Finally, field audits and traffic-conflict observations at shortlisted locations would provide engineering validation.

VIII. CONCLUSION

A data-driven framework was developed to assess reported road accident burden across 50 Indian million-plus cities using official statistics. Entropy-weighted TOPSIS, K-means clustering, sensitivity analysis and a burden–severity matrix provide complementary summaries of accident, fatality and injury counts. Delhi, Bengaluru and Jabalpur were the highest-ranked cities under the composite burden score. The five-cluster solution was stable across the tested initializations, while sensitivity analysis showed that the ranking changed more substantially when fatalities were given a dominant weight.

The principal value of the framework is a transparent screening procedure that combines multiple indicators without conflating different geographic aggregations or overstating what the data can establish. Its outputs can help direct subsequent data collection and local safety assessment, but they should not replace exposure-based risk modelling, road-safety audits or site-level engineering investigation.

A. Data Availability

The analysis is based on the publicly available MoRTH Road Accidents in India 2023 report [3]. The extracted city-level data and computational files should be deposited with the final thesis or supplementary material after the underlying transcriptions are checked against the source tables.

B. Declaration

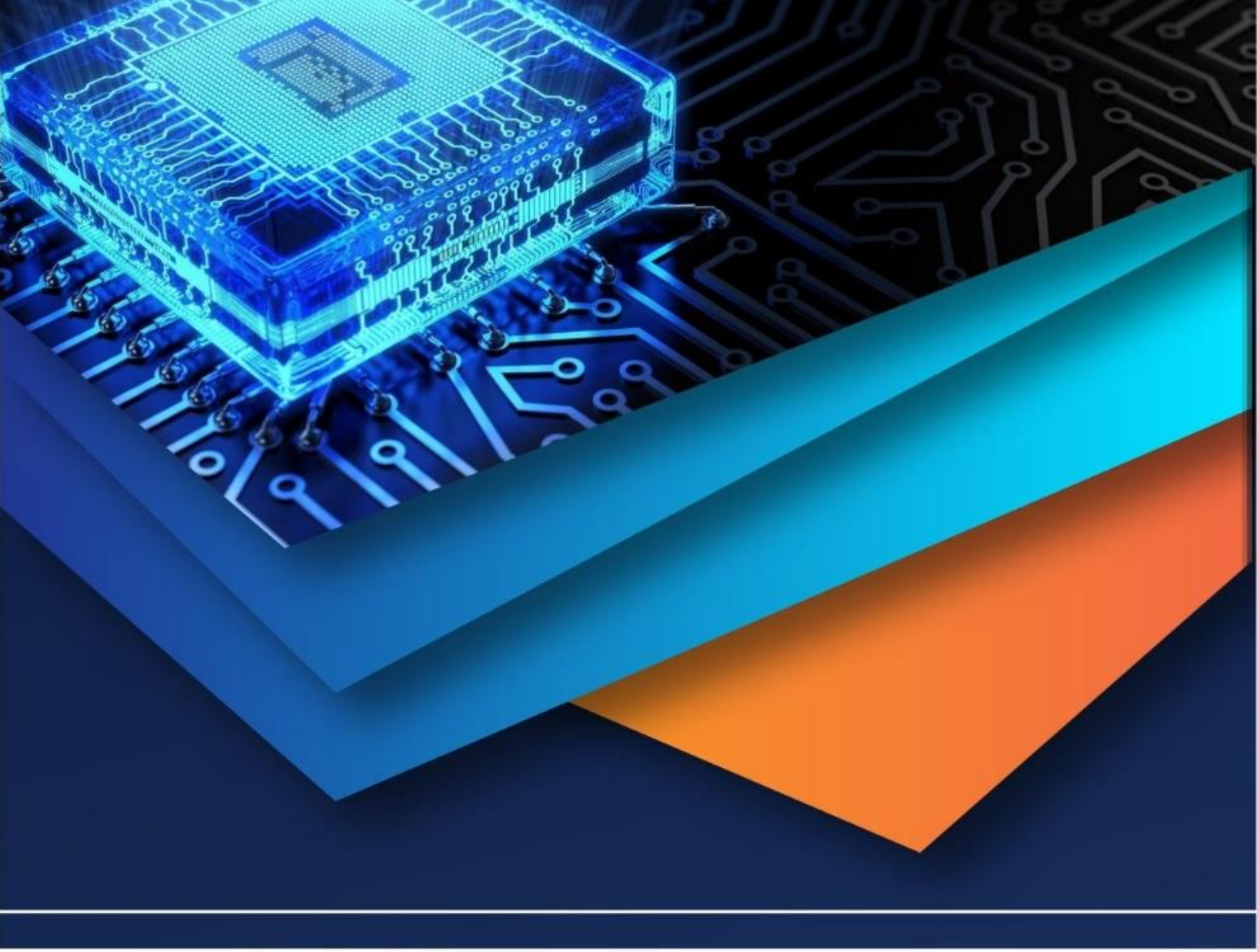
Conflict of interest: The author declares no known competing financial interests or personal relationships that could have appeared to influence the work. (Confirm and amend this statement before submission.)

REFERENCES

[1] World Health Organization. (2023). Global status report on road safety 2023. Geneva: WHO. https://www.who.int/publications/i/item/9789240086517
[2] World Health Organization. (2021). Global plan for the decade of action for road safety 2021–2030. Geneva: WHO.
[3] Ministry of Road Transport and Highways. (2025). Road accidents in India 2023. Transport Research Wing, Government of India. https://morth.nic.in/sites/default/files/Road-Accident-in-India-2023-Publications.pdf
[4] World Health Organization. (2010). Data systems: A road safety manual for decision-makers and practitioners. Geneva: WHO.
[5] Barffour, M., Gupta, S., Gururaj, G., & Hyder, A. A. (2012). Evidence-based road safety practice in India: Assessment of the adequacy of publicly available data in meeting requirements for comprehensive road safety data systems. Traffic Injury Prevention, 13(Suppl. 1), 17–23. https://doi.org/10.1080/15389588.2011.636780
[6] Mohan, D., Tiwari, G., & Mukherjee, S. (2016). Urban traffic safety assessment: A case study of six Indian cities. IATSS Research, 39(2), 95–101. https://doi.org/10.1016/j.iatssr.2016.02.001

[7] Paul, A. B. (2024). Investigating and modeling motorcycle crash rates at urban unsignalized intersections in India. Traffic Injury Prevention, 25(6), 852–859. https://doi.org/10.1080/15389588.2024.2348729
[8] Anderson, T. K. (2009). Kernel density estimation and K-means clustering to profile road accident hotspots. Accident Analysis & Prevention, 41(3), 359–364. https://doi.org/10.1016/j.aap.2008.12.014
[9] Maji, A., Velaga, N. R., & Urie, Y. (2018). Hierarchical clustering analysis framework of mutually exclusive crash causation parameters for regional road safety strategies. International Journal of Injury Control and Safety Promotion, 25(3), 257–271. https://doi.org/10.1080/17457300.2017.1416485
[10] Hwang, C. L., & Yoon, K. (1981). Multiple attribute decision making: Methods and applications. Springer.
[11] Shannon, C. E. (1948). A mathematical theory of communication. Bell System Technical Journal, 27(3), 379–423; 27(4), 623–656.
[12] MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability (Vol. 1, pp. 281–297). University of California Press.
[13] Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7
[14] Hubert, L., & Arabie, P. (1985). Comparing partitions. Journal of Classification, 2, 193–218. https://doi.org/10.1007/BF01908075
[15] Papadimitriou, E., Yannis, G., Bijleveld, F., & Cardoso, J. L. (2013). Exposure data and risk indicators for safety performance assessment in Europe. Accident Analysis & Prevention, 60, 371–383. https://doi.org/10.1016/j.aap.2013.04.040

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)