

Pedestrian Safety Assessment on Roads in Rural Areas

ABSTRACT

Pedestrian fatalities on roads in rural areas represent a critical public health concern around the globe, which is underexplored in low-and-middle-income countries (LMICs) like India. Unlike access-controlled highways in high-income countries, rural roads in India function as mixed-use corridors, accommodating high-speed vehicular movement alongside significant pedestrian activity generated by settlements, roadside landuse, and informal access points. Despite the growing burden of pedestrian fatalities on such corridors, existing safety assessments are often limited to single-method approaches and fail to capture the complex, multi-dimensional, and interaction-driven nature of pedestrian crash risk. This thesis addresses this gap by developing an integrated, multi-level analytical framework that combines geospatial analysis, epidemiological methods, data mining techniques, advanced Bayesian modeling, and machine learning approaches to comprehensively assess fatal pedestrian crashes on these high-speed rural roads and achieve five complementary objectives.

The study begins with a statewide geospatial assessment of pedestrian fatalities on rural roads in the state of Haryana using multi-source datasets, including crash records from 2017-2022 (excluding 2020 due to the COVID-19 lockdown), satellite-derived landuse, and gridded population data, along with open-source datasets from OpenStreetMap. All the roads in rural areas of these three categories, National Highways (NH), State Highways (SH), and Major District Roads (MDR), were mapped. The analysis reveals that approximately 68% of fatalities occur in rural areas, with Haryana exhibiting one of the highest fatality rates in India (~20 fatalities per 100,000 population). Pedestrian fatalities are disproportionately concentrated along NHs, where increasing population exposure, built-up expansion, and village density significantly amplify risk. The findings demonstrate that rural highways are evolving into activity corridors with intense pedestrian-vehicle interactions, thereby necessitating context-specific safety interventions.

The identification and quantification of key risk factors at a micro-level (100 m segments) influencing fatal pedestrian crashes were assessed using a matched case-control study design. This analysis was conducted on six selected road stretches (4 NHs, 1 SH, and 1 MDR) with high pedestrian fatality rates and pedestrian densities. The results indicate that fatal pedestrian crashes are primarily driven by location-specific and infrastructure-related variables rather than traffic volumes or speeds alone. Junctions, settlements, authorized median gaps, flyover ends, service roads, and roadside activity generators (such as eateries, schools, bus stops, and petrol pumps) exhibit significantly higher odds of fatal crashes, highlighting the dominant role of

access management and the roadside environment in shaping pedestrian risk. The matched design effectively controls confounding roadway characteristics, ensuring robust estimation of risk relationships.

The study further employs association rule mining to uncover co-occurrence patterns of risk factors across segments and finds that crash risk arises from the combined effects of multiple factors. The results reveal that fatal pedestrian crashes are associated with stable and recurring combinations of features, such as junctions located within settlements combined with median gaps and high exposure conditions. These findings have a direct bearing on specific interventions near junctions and at median gaps when the highway passes through a settlement.

To quantify pedestrian crash occurrence and develop predictive safety performance functions, Hierarchical Bayesian crash frequency models, including Negative Binomial, Random Parameters Negative Binomial, Negative Binomial-Lindley, and Random Parameters Negative Binomial-Lindley, were estimated using a disaggregated segment-level dataset. The results indicate that pedestrian crash data exhibit overdispersion, excess zeros, and substantial unobserved heterogeneity. Among the evaluated models, the Random Parameters Negative Binomial-Lindley model provides the best statistical fit and predictive performance, confirming the importance of simultaneously accounting for excess zeros and heterogeneity. Key variables influencing crash frequency include population exposure, roadside activity generators, junctions, and access-related features, while certain geometric and traffic-related variables exhibit context-dependent effects.

Finally, a machine learning-based classification framework was developed to categorize road segments into different classes based on the number of fatal pedestrian crashes in them. These classes are Class 0: no fatal crash, Class 1: one fatal crash, and Class 2: two or more fatal crashes. Three tree-based models, starting with a Decision Tree and progressing to ensemble-based models, i.e., Random Forest (RF) and Extreme Gradient Boosting, were developed. The RF model demonstrates superior performance, with an accuracy of 0.8787 and weighted average precision, recall, and F-score of 0.8748, 0.8787, and 0.8763, respectively. The highest area under the receiver operating characteristic curve (0.801) further confirms that this model outperforms the conventional logistic regression and other tree-based models. The application of the synthetic minority over-sampling technique for class imbalance handling significantly improves the identification of the minority high-risk class. Shapley additive explanations-based interpretability analysis assigns an importance value to each feature for each class and reveals that high-risk segments are strongly associated with junctions, high population, pedestrian

exposure, traffic intensity, roadway geometry, and roadside activity features, with nonlinear and interaction effects playing a dominant role in classification.

Overall, the findings of this thesis provide robust evidence that complex interactions among exposure, roadway design, traffic conditions, and the roadside environment govern pedestrian crash risk on rural roads. The integrated methodological framework proposed in this study advances both theoretical understanding and practical assessment of pedestrian safety by moving beyond isolated analyses toward a comprehensive, data-driven approach. The outcomes offer actionable insights into access management, infrastructure design, and targeted interventions, contributing to the development of safer rural highway systems in India and other LMICs.