Article:
Ling, C., Xie, B. and An, Z. orcid.org/0000-0003-2577-761X (2025) Traffic safety in relation to multidimensional street network and land use features: A nonlinear analysis with population heterogeneity. Applied Geography, 183. 103737. ISSN 0143-6228

https://doi.org/10.1016/j.apgeog.2025.103737

Traffic Safety in Relation to Multidimensional Street Network and Land Use Features: A Nonlinear Analysis with Population Heterogeneity

Abstract: Road traffic crashes remain a critical concern for public health and sustainable urban development. In recent years, there has been an emerging interest in applying nonlinear approaches to examine the relationship between the built environment and crash occurrence. Extending this line of inquiry, this research examines how diverse dimensions of street network configuration (geometry, hierarchy, topology) and land use (type, intensity, diversity) influence traffic crash density (i.e., crashes per unit area), and investigates how such effects vary by zonal population composition. within a nonlinear framework. Using data from Wuhan, China, and gradient boosting decision trees, we find that street topology and land use density are the most influential correlates in explaining the variation in crash density. Nonlinear associations of street network and land use characteristics with crash density are prevalent, with most variation occurring within a specific threshold range. Moreover, the effects of these built environment characteristics vary significantly across zones with differing age and income structures. Zonal elderly population density amplifies the effects of most street network and land use characteristics on crash density. Low-income zones demonstrate a greater sensitivity to changes in certain built environment features, such as street density, residential land ratio, and land use diversity, resulting in more pronounced increases in crash density. Our findings provide a more comprehensive and nuanced understanding of the links between the built environment and traffic safety, and call for both recognitional and distributive considerations of spatial justice to be incorporated into traffic safety interventions.

Keywords: population heterogeneity; street network; land use; traffic safety; gradient boosting decision tree (GBDT)

1 Introduction

Road traffic crashes have become a major threat to global public health and the development of sustainable cities (Jiang et al. 2017). Each year, approximately 1.19 million lives are lost in traffic crashes, with tens of millions more people suffering serious injuries (WHO 2023). As an increasing share of the population settles in urban areas, the need for traffic safety-oriented spatial planning becomes ever more pressing. A growing body of literature at the intersection of transport geography and planning has examined how built environmental factors are associated with the occurrence (frequency, rate, and density) of traffic crashes. Among these factors, street networks (Zhang et al. 2015, Wang et al. 2018) and land use (Xie et al. 2019, Qiao et al. 2020), which condition the emergence of travel demand, shape the spatial configuration of mobility systems, frame traffic speeds, and contribute to the formation of spatial conflicts (Saha et al. 2020, Ewing and Dumbaugh 2009), have received wide attention. This line of inquiry is driven by the recognition that modifying these factors potentially enables more proactive, wide-reaching, and enduring improvements in traffic safety through spatial planning, in contrast to passive measures such as vehicle safety technologies or occupant protection (Ewing and Dumbaugh 2009, Dumbaugh and Li 2011).

Most existing studies have relied on linear or generalized linear model to examine the association of street network and land use patterns with crash occurrence (Marshall and Garrick 2011, Dumbaugh and Zhang 2013, Chen and Shen 2016, Zheng et al. 2021). These approaches implicitly impose a (near-)linear functional form on the relationship, a presumption that is not sufficiently justified theoretically and empirically. Consequently, such models may fail to capture more complex, nonlinear, or threshold effects that likely exist, particularly in the context of diverse urban environments (Zhou et al. 2025). Methodologically, nonlinearity suggests the existence of variable thresholds, beyond which relationships may shift or even reverse (Ding et al. 2018a). For example, several recent studies have identified nonlinear and threshold effects of built environment features on travel behavior (Yang et al. 2021, Ding et al. 2018b, Shao et al. 2022). Although these studies do not directly examine traffic safety, their findings suggest that similar nonlinear mechanisms may also exist in the relationship between the built environment and crash outcomes.

Nonetheless, nonlinear analytical frameworks remain relatively underutilized in traffic safety research. While a few studies have adopted nonlinear approaches to examine the association between crash occurrence and street network or land use patterns, they have been limited in scope. Specifically, they have predominantly focused on characteristics in narrow dimensions—such as street density or land use type—rather than systematically considering the multidimensional nature of such characteristics. This limited perspective fails to capture the structural, spatial, and functional diversity inherent in these two built environmental elements, including important features such as street network geometry, hierarchy, and topology, as well as land use type, intensity, and diversity. Although some of these features have been considered—sometimes in combination—within linear analytical frameworks, there has been little systematic effort to integrate the full spectrum of these multidimensional attributes into nonlinear analyses. This omission limits our accurate understanding of the complex interactions between built environmental characteristics and crash occurrence.

Moreover, some recent studies have highlighted that the impact of the built environment on traffic safety may vary across different subpopulation groups (Barajas 2018, Lee et al. 2020b, Shin 2023, Yu et al. 2022, Pirdavani et al. 2017). For example, low-income areas are often found to experience higher rates of pedestrian crashes and fatalities than high-income areas, likely due to underlying infrastructure deficiencies such as inadequate street networks (Quistberg et al. 2022),

suboptimal land development, and imbalanced traffic organization (Stoker et al. 2015, Noland and Laham 2018). Similarly, population aging introduces further variation in the effects of the built environment on road safety, as elderly population may experience different levels of risk exposure and crash severity compared to younger groups. These differences arise not only from age-related declines in physical and cognitive abilities, but also from varying degrees of sensitivity to built environment features that may not be age-adaptive (Grise et al. 2018, Wilmut and Purcell 2022).

Despite these observations, zonal demographic and socioeconomic compositions, such as the age and income structure of the population, are still most often treated as control variables or aggregated attributes in analyses, rather than explicitly modeled as moderators that might alter the relationship between the built environment and traffic safety. As a consequence, this is equivalent to estimating the effects of built environment factors on crash occurrence for a homogeneous population, after accounting for the influence of these characteristics, rather than examining how these effects might vary across different subgroups. This limitation is not merely methodological but also raises concerns of spatial justice along both recognitional and distributive dimensions (Soja 2013, Walker 2012). Recognitionally, it reflects the analytical failure to acknowledge such heterogeneity in the first place—treating all populations as functionally equivalent, and thereby overlooking the differentiated ways in which built environments interact with social and demographic structures. Distributively, it risks concealing how a uniform built environment intervention, while appearing equal in its application, actually functions as a mechanism for the unequal distribution of safety benefits.

To advance understanding in this field, the present study uses a nonlinear analytical framework to examine (1) the effects of multidimensional street network and land use characteristics on traffic crash density (defined as the number of traffic crashes per unit area within a zone), and (2) how these effects vary across areas with different compositions of population subgroups. We consider multiple dimensions of structural, spatial, and functional diversity by incorporating a comprehensive set of street network measures, including hierarchy, geometry, and topology, as well as land use measures, including type, density, and diversity. We examine how these effects vary across zones with different income and age compositions by stratifying the population accordingly. We employ an advanced, explainable machine learning method, the gradient boosting decision tree (GBDT) model, to capture both nonlinear relationships and disparities in traffic crash density across population subgroups. Our research findings contribute to a more comprehensive and nuanced understanding of how built environment features affect traffic safety. They offer practical value by generating targeted, localized evidence to inform spatial planning strategies for traffic safety improvements. Importantly, our analytical approaches explicitly integrate considerations of spatial justice into traffic safety improvements, so that our findings can support interventions that are sensitive to the needs of areas with different population compositions and are better positioned to improve traffic safety for all.

2 Literature Review

2.1 Street network, land use, and traffic crash frequency

Street networks and land use are two fundamental aspects of the built environment that shape both the occurrence and severity of traffic crashes. Street network configuration determines how traffic flows and concentrates within an area, affecting exposure to risk, traffic speed, and the likelihood of conflict points; land use patterns, on the other hand, influence the distribution and intensity of

115 travel demand, thereby impacting where and when crashes are most likely to occur (Ewing and
116 Dumbaugh 2009). Prior research, depending on a linear analytical framework, has showed that both
117 street networks and land use play significant roles in determining crash occurrence (Saha et al. 2020,
118 Ewing and Dumbaugh 2009).
119 For urban street network characteristics, most studies have focused on the role of geometric
120 and hierarchical features, such as intersection density, block shapes, and proportion of different
121 types of segments, in shaping crash occurrence (Marshall and Garrick 2011). Along these lines,
122 several studies have found that higher street density and smaller block sizes are associated with
123 reductions in crash frequency (Rifaat et al. 2010, Zhang et al. 2015, Wang et al. 2018). These
124 associations are commonly attributed to factors such as the more even distribution of traffic,
125 enhanced precision in traffic control, and generally lower vehicle speeds, all of which can help
126 reduce the likelihood of over-speed crashes. However, other researchers have argued that increasing
127 street density can also result in overly connected networks or more complex routing patterns, which
128 may intensify spatial conflicts between road users and thus increase crash frequency (Marshall and
129 Garrick 2011). Moreover, hierarchical features of street networks are found to be closely linked with
130 crash frequency. For example, studies suggest that a higher proportion of principal arterial roads,
131 minor arterial roads, and roads without transit points is associated with increased injury crash
132 occurrence (Wier et al. 2009, Huang et al. 2010).
133 Beyond the geometric and hierarchical aspects of street networks, a few studies have
134 investigated the effect of network topology on crash frequency. Topological characteristics describe
135 the underlying spatial configuration and connectivity patterns of a street network, reflecting how
136 locations are organized and interlinked, independent of geometric distance or road hierarchy.
137 Theoretically, topological metrics capture different pathways through which street network
138 structures influence crash frequency. Closeness centrality and average shortest path length both
139 reflect the overall reachability within a street network: higher closeness and lower average shortest
140 path length indicate that locations are more directly accessible to one another. Such configurations
141 tend to facilitate higher movement flows, thereby increasing the likelihood of user interactions and
142 conflicts (Zhang et al. 2015). In contrast, betweenness centrality captures the concept of ‘choice’,
143 highlighting street segments that serve as key routes for through-movements. Segments with high
144 betweenness concentrate more traffic and, as a result, are often subject to greater exposure and a
145 higher risk of crash events (Wang et al. 2018, An et al. 2023).
146 For land use patterns, research has primarily focused on how land use composition and mixed-
147 use development influence crash occurrence. For example, studies show that higher proportions of
148 certain land use types, such as commercial and residential land, within an area are associated with
149 increased crash frequency (Xie et al. 2019, Hu et al. 2023), likely as a result of the intensified traffic
150 activity that accompanies higher concentrations of such land uses. However, the impact of land use
151 diversity on crash occurrence remains a subject of debate. Some studies report that greater land use
152 diversity is associated with lower rates of severe cyclist injuries and fatalities—for instance, a 1.0%
153 increase in diversity may correspond to a 1.43% decrease in severe crash rates (Chen and Shen
154 2016). In contrast, other research finds a positive relationship between land use diversity and crash
155 risk, suggesting that intensified, mixed-use areas with higher traffic flows may experience more
156 crashes (Miranda-Moreno et al. 2011, Chen 2015). These divergent findings may be partly explained
157 by the limitations of conventional linear regression models, which tend to capture only average
158 global effects while overlooking abrupt changes, local variations, or threshold phenomena (Chen et

al. 2018).

Due to the limitations of generalized linear models, several recent studies have sought to employ nonlinear approaches to investigate the effects of street network and land use characteristics on crash occurrence. Ding et al. (2018c) used Poisson regression tree analysis to identify nonlinear effects of road network features—such as intersection and sidewalk density—as well as land use type and diversity, on pedestrian–vehicle collision frequency. Saha and Dumbaugh (2021) examined the relationship between pedestrian crash frequency and the lengths of various types of streets (such as local roads, arterial roads, and surface streets with five or fewer lanes), as well as the number of points of interest—a proxy for land use type—using both decision tree models and generalized additive models. Their results indicated that these relationships were generally nonlinear. Luo et al. (2023) employed an extreme gradient boosting decision tree model to examine the frequency of injurious traffic crashes, finding nonlinear associations with the lengths and geometric features of street elements (e.g., sidewalks, bike lanes, and curvature), as well as land use types. However, these studies have primarily focused on a limited set of street network and land use characteristics, and have not taken into account the heterogeneity in population composition across areas.

2.2 Traffic crash frequency and population heterogeneity

Empirical research has observed close correlations between economic status and pedestrian-involved crash occurrence (Mansfield et al. 2018, Stoker et al. 2015). Individuals living in economically disadvantaged neighborhoods tend to be notably more likely to be involved in crashes compared to those in high-income areas (Li et al. 2022, Chakravarthy et al. 2010). Higher crash exposure in these areas has been attributed to factors such as increased residential density, inadequate safety infrastructure, disconnected street networks, and suboptimal land use development (Roll and McNeil 2022, Yu et al. 2022, Yu 2014, Noland et al. 2013). For example, Morency et al. (2012) found that the lack of safe pedestrian crossings at intersections in low-income neighborhoods significantly contributes to elevated pedestrian collision rates. Moreover, studies indicate that distinct risk factors contribute to crash frequency across different income groups, even after controlling for variations in neighborhood environments based on income composition (Yu 2014, Yu et al. 2022). This suggests that, beyond environmental factors, individuals living in neighborhoods with different income levels may themselves exhibit divergent travel behaviours, safety practices, and perceptions of risk, which in turn contribute to observed disparities in crash occurrence.

Several studies have highlighted inequalities in crash occurrence associated with zone-level demographic characteristics, particularly in high-density Asian cities (Ling et al. 2024). These disparities are often attributed to differences in the proportions and densities of elderly residents, which result in varying levels of vulnerability among pedestrian populations and, in turn, differing levels of crash involvement. High population concentration, combined with reduced mobility among the elderly population, poses significant challenges for implementing effective road safety improvements. Research consistently shows that elderly populations are more susceptible to injuries in traffic crashes (Feng et al. 2021) and require supportive walking environments (Grise et al. 2018). When confronted with imminent traffic conflicts, the elderly population are more likely to make erroneous decisions, due to age-related declines in visual acuity, auditory perception, and physical agility (Dommes et al. 2013, Luo et al. 2023, Niebuhr et al. 2016).

Existing traffic safety research has predominantly treated zone-level socioeconomic and demographic factors, such as income levels and age composition, as control variables capturing

contextual effects. While this approach helps to isolate the effects of built environment features, it often overlooks the complex and potentially interactive roles these population attributes play through their interactions with the built environment, thereby altering crash risks. As a result, current empirical models may oversimplify the role of zone-level socioeconomic and demographic composition, which not only independently influences traffic safety but may also moderate the effects of built environment factors on crash occurrence by either amplifying or attenuating their impacts.

3 Research design

Figure 1 illustrates the analytical framework of our research; the colored pathways highlight our focal points of analysis. Under a nonlinear analytical framework, we address two central areas of interest. First, we consider the multidimensional characteristics of street networks (hierarchy, geometry, topology) and land use (type, intensity, diversity), and examine their relationships with zonal road crash density. Following the existing literature (Ewing and Dumbaugh 2009, Dumbaugh and Li 2011), we hypothesize that such relationships are mediated by traffic volume, traffic speed, and traffic conflicts. Next, we investigate how these relationships are moderated by zone-level demographic and socioeconomic compositions, specifically zonal age composition and income levels. Finally, nonlinear explainable machine learning methods are utilized to model nuanced relationships between variables, following recent advances in the fields (Liu et al. 2024, Liu et al. 2025b).

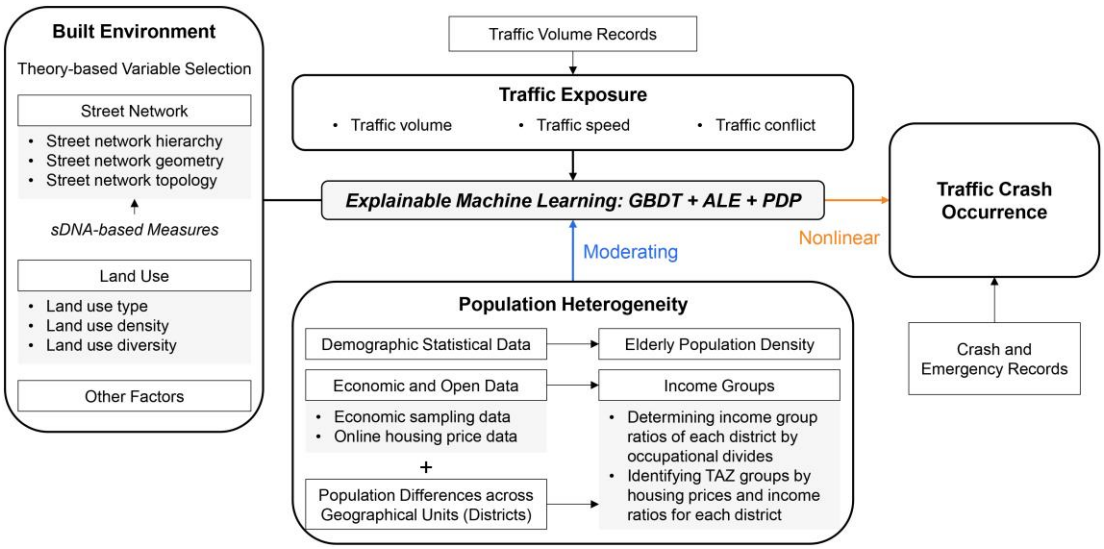


Figure 1. The analytical framework.

3.1 Case study area

This study took Wuhan City, China, as an example and used traffic analysis zones (TAZs) as the statistical units to trace the high frequency of traffic crashes and the dynamic population changes in these small regions (**Figure 2**). Severe traffic crash record data were officially collected by the Wuhan Medical Center for Emergency, wherein each record represents a traffic crash accident associated with fatality or disability. A previous study has demonstrated the good representativeness and accuracy of this dataset (Xie et al. 2019). We performed a geocoding verification step for the data and excluded records with incomplete addresses, duplicate information, or ambiguous address

names. From 2012 to 2015, a total of 17,711 severe crashes occurred in Wuhan, of which 71% were within our study area. The average annual crash rate was calculated to be 55.38 per 100K people, significantly higher than peer cities with similar population and urban forms.

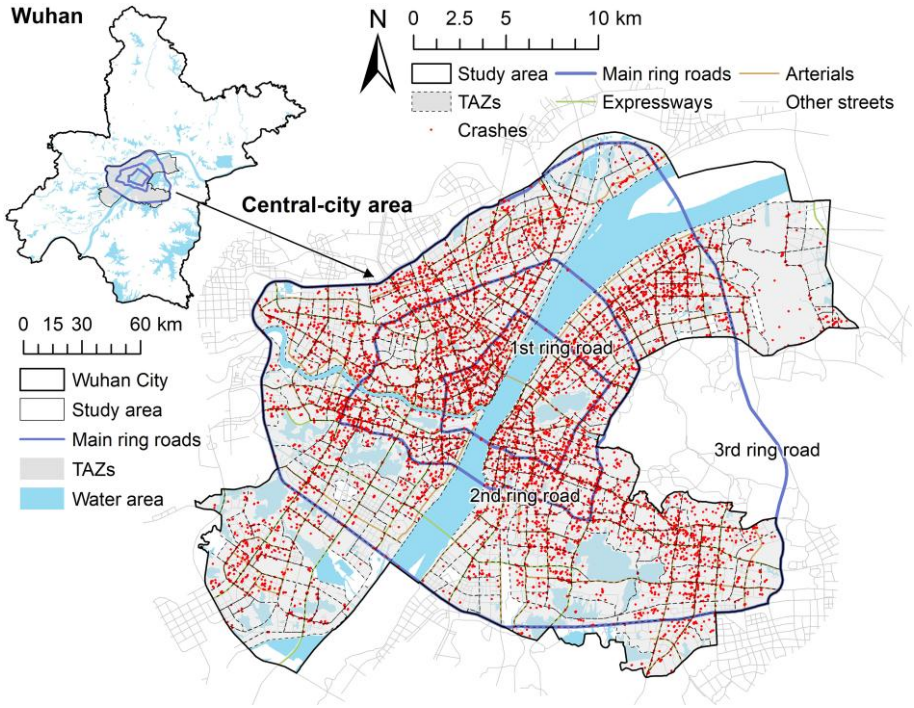


Figure 2. Traffic crashes in central Wuhan.

3.2 Variables and measurements

Table 1 presents the definitions and descriptive statistics of variables used in our study. The dependent variable is traffic crash density in each Traffic Analysis Zone (TAZ), and the independent variables include built environment as well as demographic and socioeconomic (i.e., population heterogeneity) features.

3.2.1 Traffic crash density

The crash density of a specific TAZ was defined as the ratio of severe crashes to the area of that TAZ, following Qiao et al. (2020):

$$Crash\ density = \frac{Number\ of\ traffic\ crashes}{TAZ\ land\ area} \quad (1)$$

where the TAZ land area refers to the TAZ's total land area excluding water bodies. Standardizing by area is necessary because some TAZs—particularly those at the edges of the study area—can be much larger than others. Using raw crash frequency as the dependent variable may introduce bias and uncertainty, as larger TAZs could report higher crash counts simply due to their size rather than a genuinely higher risk.

3.2.2 Traffic exposure

Following Ewing and Dumbaugh (2009) and Saha et al. (2020), we included traffic volume in the

model to account for traffic exposure levels (Figure 3(a)). Traffic volume was measured as annual average daily traffic, using vehicle count data that was automatically recorded by road cameras and provided by the Wuhan Traffic Management Bureau.

Table 1. Variable measurements and descriptive statistics.

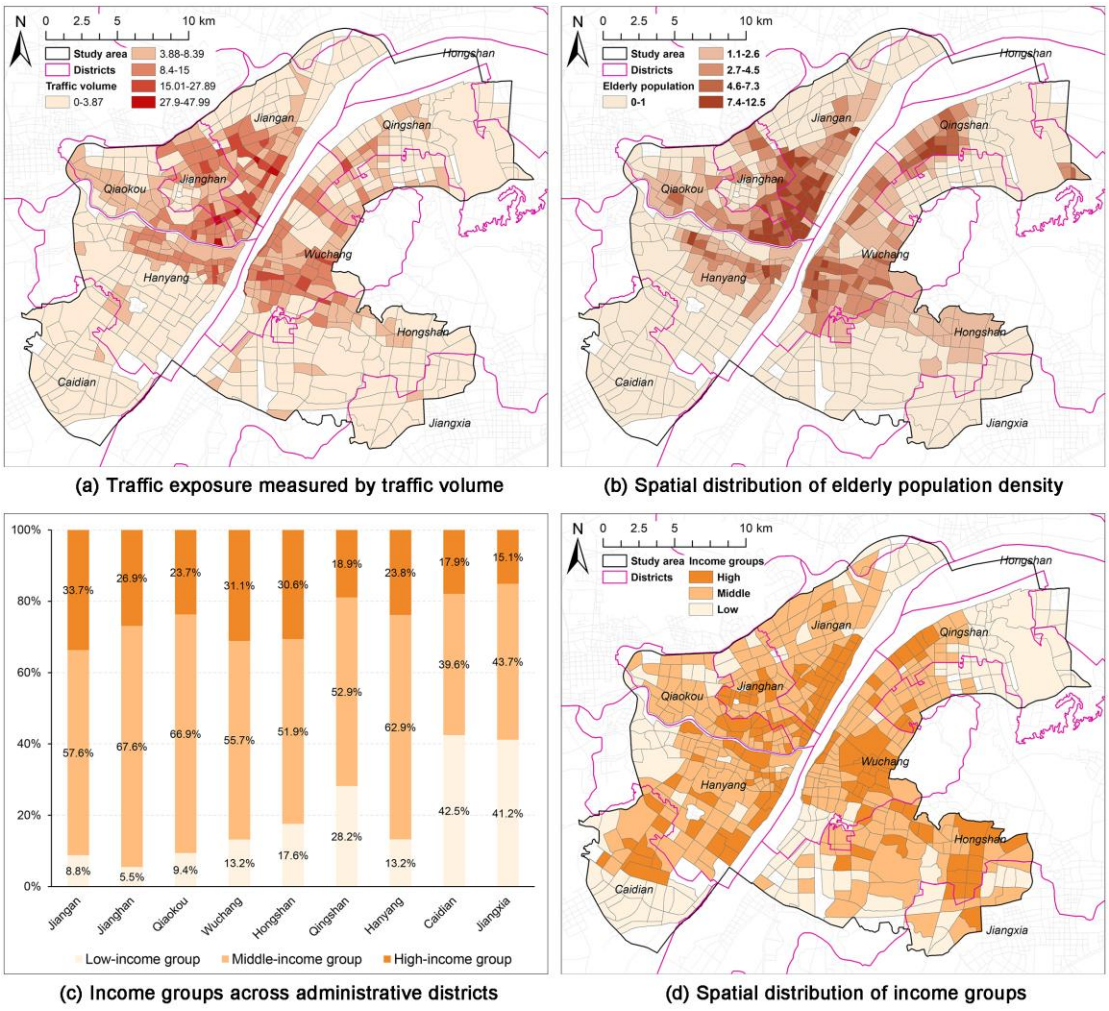
Variables	Measurements	Mean	S.D.	Min	Max
Dependent Variable					
Crash Density	Number of severe traffic crashes per TAZ (counts/km²)	34.55	28.79	0.98	174.67
Independent Variables					
Traffic Exposure					
Traffic Volume	Annual average daily traffic per TAZ (10 K/ km²)	6.05	6.12	0	47.99
Population Heterogeneity					
Population Density	Population density (K/km²)	17.56	17.49	0.04	92.92
Elderly Population Density	Population density of people over 60 years old (K/km²)	2.3	2.59	0	12.5
Income Groups	Three groups based on zonal income estimations	{1,2,3}			
Street Network					
I: Street Network Hierarchy					
Arterial Street Ratio	Arterial street ratio to total streets	0.25	0.21	0	1
Local Street Ratio	Local street ratio to total streets	0.42	0.22	0	1
II: Street Network Geometry					
Street Density	Street length per TAZ (km/km²)	7.79	2.96	0.71	20.42
Avg. Vehicle Lanes per Segment	Average number of motorways per street segment	4.01	0.8	2	7
Avg. Sidewalks per Segment	Average ratios of sidewalks (1=one side, 2=two sides)	1.61	0.46	0.13	2
III: Street Network Topology					
Street Degree Centrality	Average degree centrality per street	4.57	0.74	0.95	6.1
Street Closeness Centrality	Average closeness centrality per street	0.21	0.13	0	1
Street Betweenness Centrality	Average betweenness centrality per street	0.16	0.13	0	1
Avg. Shortest Path Length	Average shortest path length of the street network (km)	0.93	0.4	0.35	4.16
Land use					
I: Land Use Type					
Commercial Land Ratio	Ratio of commercial and business land	0.07	0.1	0	0.72
Residential Land Ratio	Ratio of residential land	0.41	0.23	0	0.96
II: Land Use Intensity					
Building Density	Building area per TAZ	0.17	0.11	0	0.48
Floor Area Ratio	Building floor area per TAZ	0.92	0.68	0	3.03
III: Land Use Diversity					
Land Use Mix	Entropy index of six-type land uses	0.68	0.14	0.11	1
Other Built Environment Factors					
Distance to CBD	Distance to city central business districts (CBDs) (km)	4.84	3.47	0.28	14.99
Bus Stop Density	Density of bus stops (counts/km²)	4.78	4.1	0	26.03

Notes: Land use mix $= -(1/\ln K)\sum_{k=1}^{K} p_k \ln(p_k)$ where p_k is the share of k-th land use type. The

index ranges from 0 to 1, with larger values pointing to more diverse land use layouts.

3.2.3 Population heterogeneity factors

To capture zonal demographic heterogeneity, we used population density and elderly population density as indicators of the demographic composition within each TAZ. In this study, the elderly population refers to individuals aged over 60 years. In 2015, the proportion of the elderly population in Wuhan was 12.37%. 221 TAZs had a population density exceeding 2,000 persons/km², while 96 TAZs had a density greater than 5,000 persons/km². Areas with higher population density were predominantly located in long-established neighborhoods and traditional downtown districts (Figure 3(b)).



(a) Traffic exposure measured by traffic volume

(b) Spatial distribution of elderly population density

(c) Income groups across administrative districts

(d) Spatial distribution of income groups

Figure 3. Traffic exposure, aged population, and income distribution in Wuhan.

Zonal socio-economic heterogeneity was represented by the stratification of zones based on their residents' income levels. We combined occupational data and housing price data to determine zonal income levels. Population and occupational data for 2015 were acquired from the official data source from the Wuhan Community Grid Center and the 1% National Population Sampling Survey, respectively. We collected a dataset consisting of 7,746 housing units with sale prices as of October

2015 from Anjuke (https://m.anjuke.com/wh/), the largest online housing purchase and rent platform in China. This data was applied as references.

Three steps were arranged to divide TAZs into three categories according to income and occupational structures in each district of Wuhan:

First, the proportions of the shares of three income levels based on zonal occupational distribution were calculated. The 2015 National 1% Population Sampling Survey data was used to compute district-level occupation features. The original 7-point occupational scale—including: 1) leaders of government, party, and organizational bodies, 2) professional and technical workers, 3) clerical and administrative staff, 4) social and personal service workers, 5) agricultural and environmental workers, 6) manufacturing workers, and 7) other unclassified workers—was reduced to three income groups following Rong and Jin (2023). The first two categories were classified as high-income groups, the next two as the middle-income group, and the last three as low-income groups. The results of this step were presented in Figure 3(c), showing significant socio-economic heterogeneity.

Second, a descending sequence of TAZs was generated by zonal housing prices. Housing prices can partially reflect household and regional income levels (Cai et al. 2020). Due to the unavailability of individual income data, we employed zonal housing prices as a proxy and used an inverse distance weight interpolation to supplement some income levels of each TAZ.

Thirdly, we chose different ratios of high-income (or low-income) units for each district, based on different income distributions identified in step 1. For example, 33.7% of TAZs were divided into high-income zones in Jiangan District, while only 26.9% of TAZs were identified to belong to high-income groups in Jianghan District.

As a result, there were 139 TAZs, 290 TAZs, and 145 TAZs categorized into high-, middle-, and low-income groups, respectively (Figure 3(d)). This approach took district-level variations into account, which is different from traditional single-faceted classifications that only use housing prices (i.e., representing income) or urban forms. Notably, although low-income areas make up only 56% of the population, the annual crash rate for residents in these areas was approximately 1.65 times higher than in high-income areas (93.88 vs 65.78 per 100,000 population).

3.2.4 Built environment features

We collected detailed street network data from the official urban planning department of Wuhan Municipality. Street networks have been normalized into single-line connections, encompassing motorways (including expressways, arterials, secondary roads, and branch roads), bicycle lanes, and sidewalks, according to the Municipal Master Plan of Wuhan (2010–2020). We measured street network structures to reflect potential traffic assignment performance by its hierarchical structure, geometric structure, and topological structure (Marshall 2004).

First, the hierarchical structure of the street network was measured by the share of street segments with different traffic capacities. We considered two levels of urban streets, including arterial streets and branch streets, based on their different design and planning characteristics.

Second, the geometric structure was measured by street densities and street geometric shapes.

Third, we utilized the Spatial Design Network Analysis (sDNA) to measure the topology of each street segment. Specifically, we took four indicators into account: degree (also known as connectivity), network closeness (network quantity penalized by distance), network betweenness, and average shortest path length. Closeness reflected the direct accessibility and central location of

nodes in the network. Betweenness indicated the number of geodesics passing through a node, functioning like a "bridge." Relevant technical details can be found in Cooper and Chiaradia (2020). The average shortest path length of a network measured the local network efficiency in terms of shortcut routes. The specific mathematical definitions are specified in Appendix A.

For land-use patterns, we focused on land use for trip generation functions and included land use types, land use density, and land use diversity in our analysis. Land use data was also obtained from the official parcel-level land use mapping from the 2015 Wuhan Land Use Survey. Detailed building information was retrieved from Baidu Map big data in 2015. Specifically, our analysis included six categories of land: administrative land, commercial/business land, residential land, manufacturing/industrial land, green space/park land, and others. Land use density was measured by TAZ-level building density and floor area ratio. The cross entropy index was used to quantify land use diversity (Xie et al. 2019). A higher entropy value indicates greater land use mix, while a value near zero suggests only a single land type.

For robust modeling, we conducted a thorough examination of multicollinearity among the variables and found that all independent variables had variance inflation factors below 10, indicating no problematic multicollinearity. We also explored other built environment variables such as street intersection density, street network efficiency, street network clusters, and street curvature; however, they were excluded due to significant multicollinearity or measurement redundancy.

3.3 Modeling approach

To examine the nonlinear relations between crash density and our focused variables, we applied explainable machine learning (XML) methods to estimate both an overall model and separate models for high-, middle-, and low-income TAZs. While advanced econometric models offer clear advantages for addressing certain methodological problems (such as spatial dependency and hierarchical data structure) in traffic safety analysis, machine learning approaches provide unique and complementary strengths. Notably, machine learning models can automatically capture complex, nonlinear, interactive patterns in the data, without requiring the researcher to pre-specify functional forms based on prior theoretical knowledge. In contrast, econometric models rely on polynomial functions or other transformations to address nonlinearity, a process that becomes increasingly complex and difficult to interpret as higher-order terms are added. By flexibly uncovering intricate relationships and interactions, XML methods offer additional insights that are often difficult to achieve using econometric approaches.

The Gradient Boosting Decision Tree (GBDT) model is an advanced machine learning technique that combines decision trees with a boosting algorithm to effectively capture nonlinear effects (Friedman 2001). GBDT offers several key strengths: (1) high predictive accuracy and computational efficiency; (2) the capacity to handle diverse types of independent variables; (3) effective management of multicollinearity; and (4) no requirement for specific distributional assumptions (Yang et al. 2021). Owing to its robust predictive performance, efficiency, and strong ability to model complex nonlinear relationships, GBDT has been widely applied in safety science and transportation research (Ding et al. 2019, Liu et al. 2020, Shao et al. 2022, Yang et al. 2020).

The specific algorithm process and technical details of the GBDT have been well-documented in prior studies (Ding et al. 2018a, Yang et al. 2021). In this study, we provide a concise overview of the model (Appendix A). In the model fitting steps, three important parameters were regularized using the grid search method: shrinkage rate, maximum estimators, and the tree depth parameter.

Following previous studies (Ding et al. 2018b, Yang et al. 2021), we set the shrinkage rate to 0.001. We tested three optional maximum tree numbers in the set {3,000, 5,000, 10,000} and found that 5,000 was suitable for the overall model, while 3,000 was appropriate for grouped models, balancing the risk of overfitting and underfitting. Tree depth (also known as complexity) determines the interaction among explanatory variables. The existing literature suggests that tree depths in travel behavior models typically range from 5 to 15 (Tao and Næss 2022, An et al. 2022). We explored depths from 1 to 20. Additionally, five-fold cross-validation was used for performance evaluation based on root mean square error.

To interpret the black box of machine learning, we quantified relative importance (RI), accumulated local effect (ALE), and partial dependence plot (PDP) to understand the relationship between independent variables and predicted values. RI captures the relative contributions of independent variables to predicting and explaining dependent variables, similar to standardized beta coefficients (in Section 4.1). ALE is a microscopic method reflecting partial influence and effectively managing multicollinearity (Tao and Næss 2022). This method was used to illustrate nonlinear effects of street networks and land use on crash density along the same axis for direct comparisons (in Section 4.2). ALE plots demonstrate the marginal effect of an independent variable on the dependent variable while controlling for other independent variables. The impact of the built environment on crash density varied across its range, indicating nonlinear effects. Two-dimensional PDPs were employed to capture interactions with other independent variables such as socioeconomics and the built environment. These interactions revealed effects larger or smaller than the sum of individual effects of the two variables (in Section 4.3). Technical details about the algorithms can be found in Appendix A, Parts 3 and 4.

To compare model performance, we used pseudo R^2 values, Root Mean Squared Error (RMSE), and Relative Absolute Error (RAE) to evaluate and compare the GBDT model with the commonly used multiple linear regression and geographically weighted regression (GWR) models.

4 Results and Discussion

Through grid parameter search with five-fold cross-validation, the optimal tree depths for the four GBDT models were 15, 9, 10, and 8, respectively. In other words, each optimal model represents the result of five models with the same parameters or a total of 315 (=3×21×5) models with different parameters. The best performance was achieved after 2,956, 2,153, 2,760, and 2,356 iterations for each model, respectively. The GBDT models exhibited higher pseudo R^2 values and lower RMSE and RAE values compared to the corresponding multiple linear regression models, indicating better overall performance (see Appendix B Table B1). While the results of Global Moran's I indicate some spatial autocorrelation in crash density across TAZs, the GWR models did not substantially improve pseudo R^2 values over multiple linear regression (with the largest improvements less than 0.11). For the same modeling scenarios, GBDT outperformed GWR in all three evaluation metrics. Together, these results demonstrate the clear advantages and robustness of the GBDT model for analyzing crash density patterns in our dataset.

4.1 Relative importance for explaining crash density

Table 2 presents the relative importance of variables, showing that traffic exposure, floor area ratio, and population density are the most influential variables across all models, which aligns well with

the results from previous studies on traffic crash frequencies (Wu et al. 2020, Luo et al. 2023). Moreover, the traffic volume's importance consistently ranked within the top seven across all four models, notably achieving second importance with 13.67% in high-income areas.

The prediction share of street network and land use accounted for approximately 60% across the four models. In the overall model, street network topology demonstrated the highest explanatory power at 17.56%, followed by land use density (13.34%) and street network geometry (10.9%). Among the five most significant built environment variables, floor area ratio exhibited a predictive ability of 9.86%, followed by distance to CBDs (6.85%), bus stop density (6.5%), street density (5.7%), and street average shortest path length (5.64%).

It is notable that elderly population density is an influential factor affecting traffic density in space, as it shows a strong coincidence between denser elderly populations and denser traffic crashes. This phenomenon cannot be fully captured by the population density variable since elderly population density still achieves at least 1.57% to at most 10.68% of relative importance even though total population is considered. Across income groups, floor area ratio demonstrated superior predictive ability at 15.46% for high-income zones. Street betweenness centrality exhibited the highest predictive ability in the middle-income areas, while the residential land ratio showed the highest predictive ability in the low-income areas. These observations partially confirmed the necessity of considering the status of disaggregated people stratification for traffic safety analysis.

Table 2. Relative importance of variables in the four-income models.

	Overall			**High-income**			**Middle-income**			**Low-income**		
Variables	RI (%)	Rank	Sum (%)	RI (%)	Rank	Sum (%)	RI (%)	Rank	Sum (%)	RI (%)	Rank	Sum (%)
Traffic Exposure			7.98			13.67			7.12			6.61
Traffic Volume	7.98	3		13.67	2		7.12	5		6.61	7	
Population Heterogeneity			18.58			10.53			17.50			20.69
Population Density	10.68	1		8.95	4		9.75	1		10.61	3	
Elderly Population Density	7.90	4		1.57	18		7.75	4		10.08	4	
Street Network												
I: Street Network Hierarchy			6.02			4.12			6.42			5.65
Arterial Street Ratio	3.09	16		2.54	15		2.85	18		3.52	11	
Local Street Ratio	2.94	18		1.58	17		3.57	14		2.13	15	
II: Street Network Geometry			10.90			11.13			10.04			12.38
Street Density	5.70	7		6.53	5		3.71	12		8.13	5	
Avg. Vehicle Lanes per Segment	3.37	14		3.12	11		3.19	16		2.39	13	
Avg. Sidewalks per Segment	1.83	19		1.48	19		3.14	17		1.87	16	
III: Street Network Topology			17.56			16.55			19.06			15.99
Street Degree Centrality	3.54	12		2.86	14		3.86	11		2.90	12	
Street Closeness Centrality	3.01	17		4.28	8		2.47	19		5.87	8	
Street Betweenness Centrality	5.36	9		3.76	9		8.80	2		2.19	14	
Avg. Shortest Path Length	5.64	8		5.65	6		3.93	10		5.03	9	
Land use												
I: Land Use Type			8.37			6.34			9.95			15.26
Commercial Land Ratio	3.29	15		2.90	13		3.53	15		3.71	10	

Residential Land Ratio	5.08	10		3.44	10		6.42	7		11.55	1	
II: Land Use Density			13.34			17.44			13.07			13.01
Building Density	3.48	13		1.98	16		7.79	3		1.67	18	
Floor Area Ratio	9.86	2		15.46	1		5.28	9		11.34	2	
III: Land Use Diversity			3.89			2.94			3.58			1.83
Land Use Mix	3.89	11		2.94	12		3.58	13		1.83	17	
Other Built Environment Factors			13.36			17.28			13.27			8.57
Distance to CBD	6.85	5		12.64	3		6.29	8		1.24	19	
Bus Stop Density	6.50	6		4.64	7		6.98	6		7.33	6	

Notes: The relative importance (RI) of variables captures their partial influence on the crash density and measures to what extent the variable is important to predict the relationship.

4.2 Nonlinear effects

4.2.1 Traffic exposure

Several key factors were found to exhibit evident nonlinear effects. As illustrated in **Figure 4**, the relationship between traffic volume and crash density was positively and nonlinearly correlated when the traffic volume was below 120 K vehicles per day, after which it plateaued. Moreover, when the traffic volume exceeded 100 K vehicles per day, the positive impact of traffic volume and crash density was more pronounced in high-income areas than in other areas.

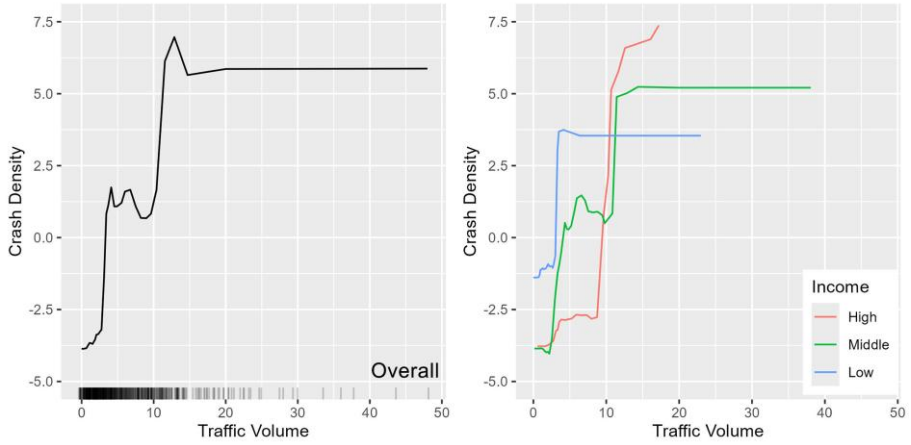


Figure 4. Nonlinear effects of traffic exposure.

Notes: The lines portray the local relationships (accumulated local effect, ALE) between independent variables and crash density, which are confined to the same *y*-axis scale. The rug plot on *x*-axis illustrates the data distribution.

4.2.2 Population heterogeneity factors

Across all income groups, increasing population density up to about 25 units is associated with a sharp increase in predicted crash density (Figure 5). Beyond that point, the marginal effect levels off, indicating a saturation effect—further increases in population density have little to no added effect on predicted crash density. Low-income areas (blue) experience the strongest positive effect of population density on crash density—even after saturation, the predicted crash density remains

consistently higher than for other groups. The ALE curves are generally nonlinear, especially at low population densities where sharp increases occur.

For elderly population density, the turning point occurs around 4,000 people per square kilometer. Before this point, a higher density of elderly population was associated with more traffic crashes. Beyond this threshold, the effect becomes less pronounced. Among the three groups, low-income areas have the highest ALE values throughout, indicating a poorer safety environment. In high-income groups (red), elderly population density appears less sensitive to traffic crash occurrence or crowding. Limited mobility resources, such as lower private car ownership and greater dependence on walking and electric bikes (especially in some East Asian cities), may further increase crash exposure for low-income populations (Roll and McNeil 2022). These characteristics imply an imbalanced safety outcome across population groups.

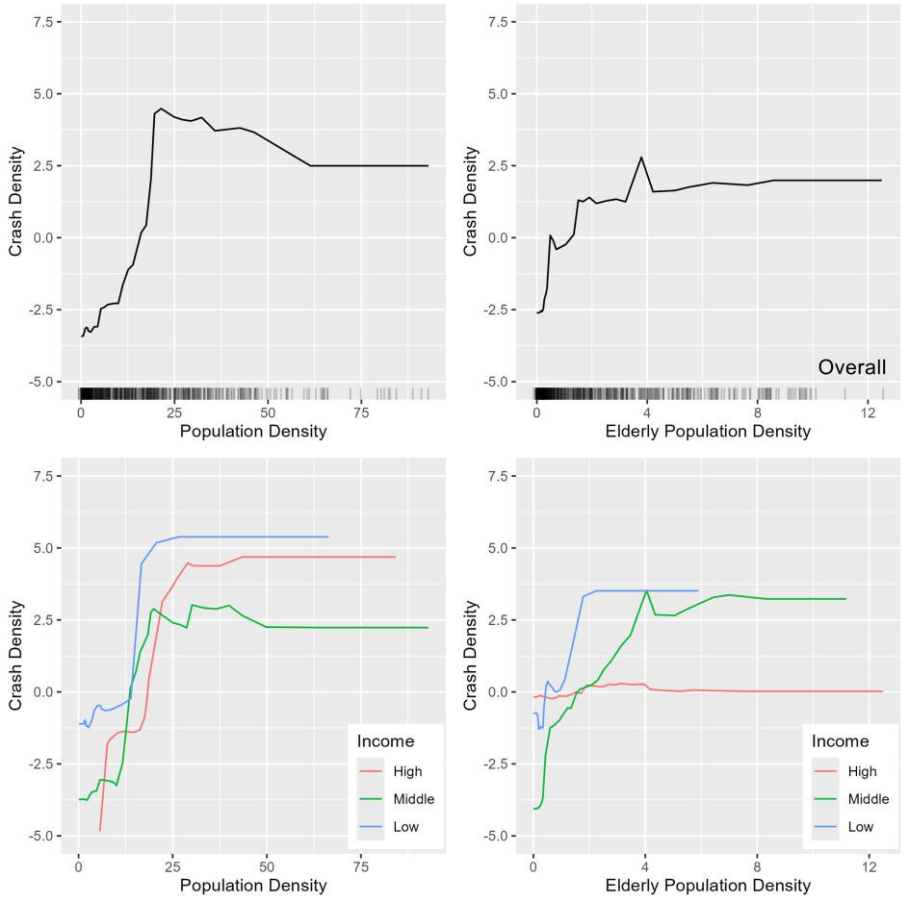


Figure 5. Nonlinear effects across population groups.

4.2.3 Street network factors

Street network, as a predominant driver, demonstrates vivid interplay with traffic crash density. **Figure 6** illustrates the significant nonlinear impact of three street network variables. In the overall model, street density shows a positive correlation with crash density within the range of 5–10 km/km^2. However, when the street density exceeds this range, the correlation slightly decreases and stabilizes, indicating a threshold effect. Moreover, in low-income areas where street density is greater than 10 km/km^2, the ALE is significantly higher than other combinations. The positive relationship is consistent with studies conducted in Florida, USA (Xu et al. 2014) and Shanghai,

China (Wang et al. 2016). This finding can be attributed to the high density of street intersections in our study area of Wuhan (Wang and Huang 2016), where the number of intersections is proportional to the total street length (r=0.91), thereby increasing the risk of traffic crashes. Furthermore, we observed a significant reduction in crash density only when the street density exceeded 10 km/km^2. A potential explanation is that a higher street density may promote low-speed travel and equitable distribution of traffic flow (Xie et al. 2022), thereby potentially encouraging the implementation of safety countermeasures (Zhang et al. 2015).

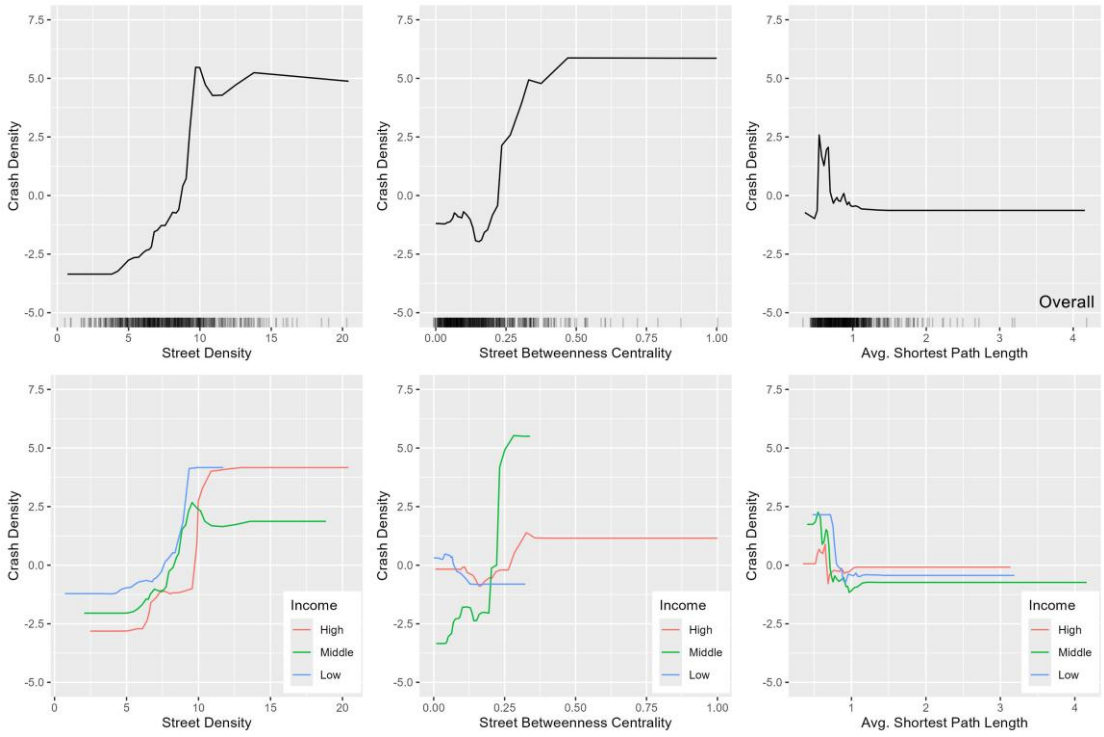


Figure 6. Impacts of street network factors.

Regarding betweenness centrality of the street network, in the overall model, the impact of betweenness on crash density is unclear below 0.2. However, once the betweenness value surpasses 0.2, even minor changes can lead to a substantial increase in crash density. Conversely, this trend does not apply to the low-income model, where an inversed negative relationship is observed. The different performances of low-income groups may be attributed to their inherent low transport connectivity and ongoing street construction activities. Additionally, the average shortest path length demonstrates a negative effect ranging from 0.5 to 1, with low-income areas showing a significantly stronger impact than the other two areas. Notably, consistent inflection points are observed on the curves across all four models, particularly at 1 km where the effect is negligible. These findings regarding other street network elements align closely with existing studies. Moreover, average shortest path length, which measures the efficiency of street networks based on Euclidean distance, exhibits a negative association with traffic crashes. This finding is consistent with previous research indicating that an increase in detours leads to a reduction in crash density per square kilometer (Zhang et al. 2015).

High-income areas, benefiting from greater motivation and substantial financial resources, are more capable of implementing comprehensive improvements to road infrastructure, with a

particular emphasis on mitigating conflicts or interactions with low-speed travelers (Guerra et al. 2022). Middle-income areas generally adhere to the overall trends observed in most variables; however, they exhibit lower sensitivity to street density while showing stronger correlations with betweenness centrality and building density. This phenomenon can be attributed to the spatial imbalance in land use and street network development between low-income and high-income areas (Wang et al. 2018, Liu et al. 2025a). Even when street density is similar, disparities exist in the integrity and continuity of the street network. For example, the proportion of sidewalks in low-income areas is approximately 70% of that in high-income areas, while the average shortest path length variable stands at 1.3. Some studies suggest that these areas often lack sustainable budgets for maintaining high-level services (Yuan and Wang 2021). Furthermore, street improvements in low-income areas are often associated with other risk factors that have a more pronounced impact, particularly at intersections. These factors can sometimes lead to fatal outcomes when conflicts occur (Yu et al. 2022).

4.2.4 Land use factors

Regarding land use types, the relationships are clear. First, an increase in residential land is linked to a significant rise in crash density in both the overall and subgroup models (Figure 7). In low-income areas, the effect of residential land on crash density is sometimes stronger than that of other variables, showing an almost linear increase when the proportion of residential land is between 0.15 and 0.7. When the share of commercial and business land is between 0.05 and 0.15, crash density reaches its highest point and is significantly affected in all three income groups (Appendix C Figure C1). Second, both floor area ratio and building density show positive nonlinear effects on crash density, but floor area ratio generally has a stronger impact. In the 0.1–0.3 range, increasing building density leads to a nearly linear increase in crash density. For floor area ratio, in low-income areas, the effect is greatest in the 0.5–1 range, while in high-income areas, the strongest effect occurs when floor area ratio is above 1. For land use diversity, the effect on crash density is minimal when the land use mix index is below 0.5. A decrease in crash density is observed when the index is between 0.5 and 0.7. However, when land use diversity exceeds 0.7, crash density rises sharply, especially in low-income areas.

Regarding land use, the positive associations between building density, floor area ratio, and crash density are consistent with previous studies (Graham and Glaister 2003, Ding et al. 2018c). Previous studies have consistently acknowledged a negative association between land use mix and crash density (Ding et al. 2018c, An et al. 2022). However, our study reveals that a moderate land use mix within the 0.5–0.7 range is associated with reduced crash risks; when the land use mix exceeds 0.7, there is a significant increase in crash density. Therefore, it is crucial to avoid indiscriminately increasing land use diversity, as it may lead to higher pedestrian density (Yu 2014) and more complex traffic flows (Xie et al. 2019), both of which can elevate crash risks.

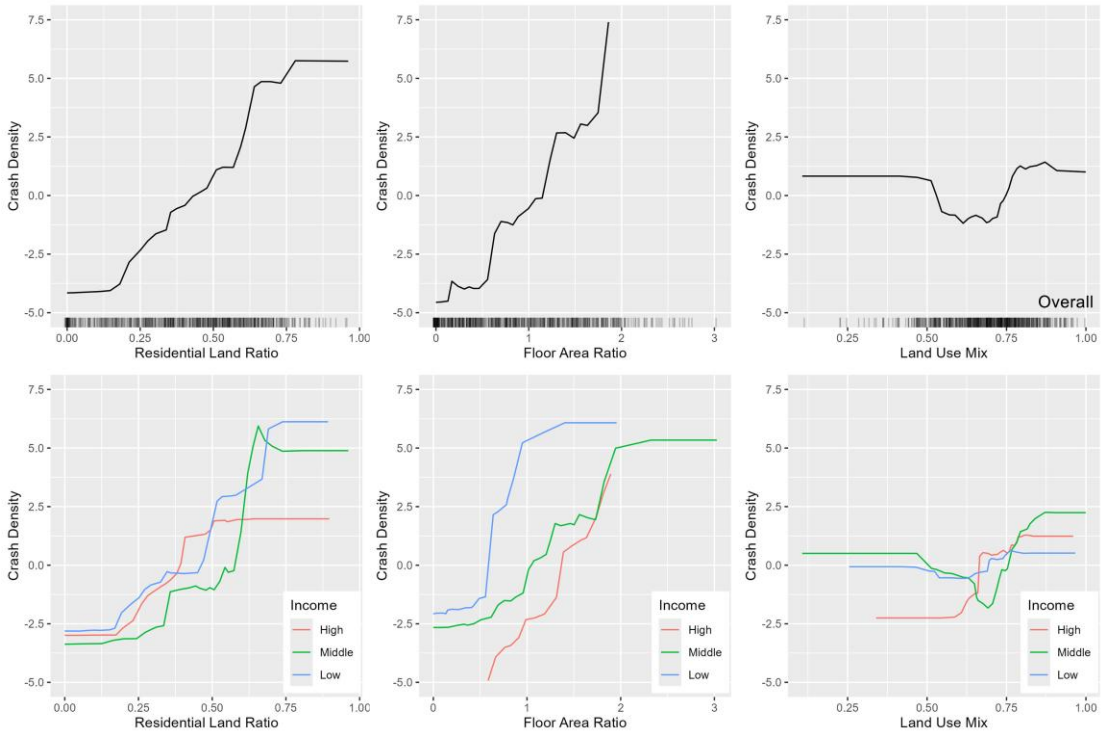


Figure 7. Impacts of land use factors.

In contrast, the low-income areas demonstrate heightened sensitivity to the residential ratio and floor area ratio. These regions are often characterized by inadequate infrastructure, largely due to historical patterns of the land use development (Noland and Laham 2018). Consequently, this has resulted in suboptimal land use and socially vulnerable neighborhoods (Yuan and Wang 2021). In Wuhan, low-income areas exhibit a residential ratio as low as 34%, accompanied by a corresponding low floor area ratio. Moreover, street density in these areas is only 71% of that in high-income areas, indicating significant delays in street construction.

4.3 Interaction effects

The above sections show strong positive associations between elderly population density and crash density, indicating that an increase in the aged population is expected to significantly raise traffic crash density. Next, we test whether the combined effect of two variables (elderly density and another variable) on crash density is larger or smaller than the sum of their individual effects. This helps us examine the interaction effects between variables. **Figure 8** shows the absolute changes in predicted crash density when elderly population density increases from 1,000 to 4,000 people per square kilometer and the other variable also increases at the same time.

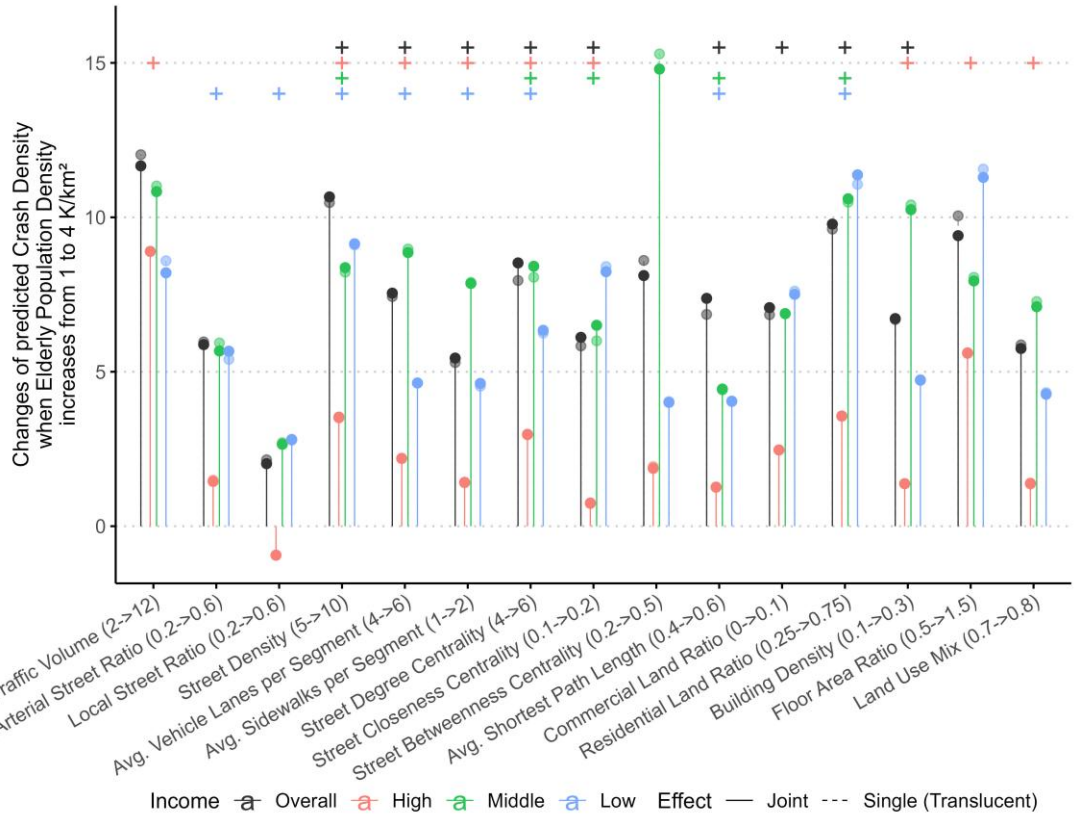


Figure 8. Interaction effects between elderly population density and other variables.

Notes: The joint effect shows the combined impact of elderly population density and each built environment factor on crash density. The single effect is the sum of their individual partial dependence effects. A colored plus sign ("+") means the joint effect is greater than the single effect, suggesting there may be an interaction between the two variables.

Among the built environment predictors, the combined effects of street density and elderly population density, as well as street degree centrality and elderly population density, on crash density exceed the sum of their individual effects, independent of zonal income levels. In other words, these two features exhibit amplified synergistic effects with elderly population density. The strongest effect is observed when street density increases from 5 km/km² to 10 km/km² and elderly population density rises from 1,000 to 4,000 people per square kilometer. In this range, crash density increases by 10.66, 3.53, 8.37, and 9.14 in the four income groups, respectively—all of which are greater than the sums of the individual effects in corresponding groups (9.48, 3.5, 8.22, and 9.11). Notably, the increase in crash density associated with rising street density is smaller in middle- and high-income areas compared to low-income areas. A similar amplified synergistic effect is observed for street degree centrality. Although the combined effects of street degree centrality and elderly density on crash density are smaller in absolute terms compared to those of street density and elderly density, the magnitude of their synergy—that is, the extent to which the joint effect exceeds the sum of individual effects—is even greater.

Amplified synergistic effects with elderly population density are also observed for several street network features, including the number of vehicle lanes, sidewalk presence, and closeness centrality. For land use, residential land ratio, commercial land ratio, and building density

demonstrate greater-than-additive effects when combined with elderly population density.

These results suggest that increases in elderly population density tend to amplify the effects of built environment features that have independent negative impacts on traffic safety. The amplification may be explained by the heightened vulnerability of elderly individuals to traffic risks arising from specific built environment conditions. Age-related declines in hearing, visual perception, motion sensitivity, and the ability to estimate contact time can compromise older individuals' capacity to respond effectively in dynamic traffic environments (Wilmut and Purcell 2022). These limitations often result in slower movement speeds, reduced situational awareness, and prolonged exposure to potential hazards (Lee et al. 2020a). Moreover, difficulties in distinguishing critical information from irrelevant cues can impair judgement during hazardous situations, particularly in complex street environment (Dommes et al. 2013, Saha et al. 2020).

Given these vulnerabilities, certain built environment features are likely to pose greater risks for elderly individuals. Specifically, features such as residential land ratio and street density are independently associated with elevated crash risk. A high residential land ratio typically leads to increased pedestrian activity and street-level interactions, while greater street density results in more frequent intersections and complex navigation demands. When these risk factors coincide with high elderly population density, older individuals—already constrained by age-related declines in perception, cognition, and mobility—are confronted with both more frequent opportunities for traffic conflict and greater difficulty in making safe navigation decisions. Consequently, the combined presence of these built environment characteristics and a high concentration of elderly residents can produce a synergistic amplification of crash risk that exceeds the impact of each factor considered independently.

5 Conclusion and Implications

This research uses a nonlinear analytical framework to examine the effects of multidimensional street network and land use characteristics on traffic crash density, and how these effects vary across areas with different compositions of population subgroups.

Our analyses revealed significant nonlinear effects of street density, street network betweenness, street average shortest path length, floor area ratio, and land-use diversity as the primary influencers on crash density. Notably, street topology exhibited superior predictive capabilities for crash density compared to other variables. In contrast to studies assuming linear relationships, our study quantifies the relative contributions of various dimensions of street networks and land use patterns to crash occurrence under more accurate, nonlinear conditions. Unlike research focused solely on nonlinear relationships, our study considers a broader range of dimensions in capturing street networks and land use characteristics. These contributions together offer clear planning implications for traffic safety improvements.

First, our results highlight which factors should be prioritized in urban planning for traffic safety. Street density, network betweenness, and land use diversity emerged as key influencers, suggesting that planners should pay close attention to these dimensions when designing safer urban environments. For instance, interventions could include optimizing street layouts to balance density and connectivity, and promoting a diverse mix of land uses to distribute travel demand more safely across the network. It is important to note that changes in street density, network betweenness, and land use diversity are associated with changes in crash density only within certain intervals. Therefore, targeted adjustments should be made according to the specific range of these variables

in a given area. Second, our findings emphasize the crucial role of street topology—a factor often overlooked in practical planning. Urban planners and engineers should adopt a holistic perspective and, before redesigning or constructing new streets, carefully consider the topological relationships between proposed and existing streets. Evaluating how overall changes in street network topology within an area may influence traffic safety outcomes can inform better design decisions and help mitigate crash risks.

We found that the influence of the built environment on traffic safety varies across zones with different socioeconomic and demographic compositions. Specifically, zone-level income and age composition moderate the association between built environment elements and crash density. Zonal elderly population density amplifies the effects of most street network and land use characteristics on crash density. Low-income zones demonstrate a greater sensitivity to changes in certain built environment features, such as street density, residential land ratio, and land use diversity, resulting in steeper and more pronounced increases in crash density. For features like building density and street network betweenness centrality, income level shapes the nonlinear relationship with crash density, with different inflection points and effect magnitudes observed across income groups.

Our findings suggest that uniform, one-size-fits-all interventions in the built environment are unlikely to yield equitable improvements in traffic safety. When population heterogeneity is ignored, policy and design decisions risk maintaining a superficial improvement in overall safety, while continuing to conceal deep-rooted disadvantages for specific groups. In terms of spatial justice, this is both a distributive issue—concerning how risks and benefits are unevenly allocated—and a matter of recognition, as it systematically disregards the specific ways in which different populations are made vulnerable or protected through the built environment. Therefore, advancing traffic safety requires integrating both recognition and distribution as equally important principles in planning practice. This involves acknowledging which groups are most vulnerable to traffic risks and, at the same time, using empirical evidence to reveal how the impacts of built environment interventions are distributed across different population segments. Planning and policy should be guided by both perspectives, ensuring that interventions are sensitive to local vulnerabilities and attentive to the unequal allocation of risks and benefits. By explicitly capturing nonlinear effects and population heterogeneity, our analytical framework supports this approach, offering practical guidance for developing strategies that are both effective and equitable in promoting urban traffic safety.

We close by outlining the main limitations of this study. First, our crash data do not allow us to distinguish between specific severity levels or transport modes involved (for example, whether a crash involved pedestrians or only motor vehicles). While aggregating these crashes can still offer meaningful planning and policy implications, analyzing crashes by severity and type could further help reveal heterogeneous risk patterns and underlying mechanisms across different crash circumstances. Second, we used crash data from 2012 to 2015. While this period was chosen to ensure temporal consistency across all variables and to maximize data quality and completeness, the use of older data may limit the generalizability of our findings to present-day practices. Third, due to data limitations, we may have omitted some variables that could influence crash density, despite the wide range of variables considered in our analysis. For example, detailed street design elements, including zebra crossings, safety islands, turnabouts, signal lights, signs, and markings, are crucial for traffic safety. However, raw datasets containing this information were unavailable. This omission may have influenced the results of our analysis by introducing endogeneity issues.

Appendix A

Part 1: Calculation methods of street network topology

(1) Degree centrality

The sDNA constructs streets as a network with the street segment as a node and intersections as edges. Degree centrality means the number of directly connected street segments.

(2) Closeness centrality

Street network closeness centrality is measured by network quantity penalized by distance (in a gravity model). The expression can be written as

$$\text{Closeness}(x) = \sum_{y \in R_x} \frac{W(y)P(y)}{d_M(x,y)} \qquad (1)$$

where x, y, and z are street segments in the network. R_x means the set of street segments in the network radius from link x. The distance according to a metric M, along a geodesic defined by M, between an origin street segment x and a destination street segment y is denoted $d_M(x,y)$. Weight of a street segment y is denoted $W(y)$. The proportion of any street segment y within the radius is denoted $P(y)$.

(3) Betweenness centrality

Betweenness counts the number of geodesic paths that pass through a street segment, i.e., the number of times the segment lies on the shortest path between other pairs of street segments.

$$\text{Betweenness}(x) = \sum_{y \in N} \sum_{z \in R_y} W(y)W(z)P(z)OD(y,z,x) \qquad (2)$$

$$\text{OD}(y,z,x) = \begin{cases} 1, \text{if x is on the first geodesic found from y to z} \\ \frac{1}{2}, y \neq z, x = y \text{ or } z \\ \frac{1}{3}, x = y = z \\ 0, \text{otherwise} \end{cases} \qquad (3)$$

where the set of street segments in the global spatial system is denoted N. OD(y,z,x) means the number of geodesic lines passing through the street segment x between y and z. The contributions 1/2 of OD(y,z,x) to reflect the end street segments of geodesics which are traversed half as often on average, as journeys begin and end in the link center on average. The contributions of 1/3 represent origin self-betweenness.

We calculate the average closeness and betweenness centrality of all street segments within each zone, and then normalize both indices to the [0,1] interval.

(4) Average shortest path length

This indicator is a local network measure representing the average shortest path length between any two street segments within the local street network (in this study, within each TAZ). The calculation is performed using the standard Dijkstra algorithm.

Part 2: Algorithmic details of Gradient Boosting Decision Tree (GBDT)

The computation and optimization of the Gradient Boosting Decision Tree (GBDT) model involve defining an objective function and employing minimization procedures. Let x denote a vector of

predictors (i.e., built environment attributes) and *f(x)* represent the approximation function of the dependent variable y (i.e., crash density). The ensemble model can be expressed as:

$$f(x) = \sum_{m=1}^{M} \alpha_m h(x;\ \theta_m) \qquad (4)$$

where $h(x;\ \theta_m)$ denotes the regression tree fitted at the m-th iteration with structure parameters θ_m, and α_m is a scalar weight that minimizes the loss function in the direction of $h(x;\ \theta_m)$.

The function *f(x)* is updated in the negative gradient direction of the loss function as follows:

$$f_m(x) = f_{m-1}(x) + \alpha_m h(x;\ \theta_m) \qquad (5)$$

where α_m is obtained by solving:

$$\alpha_m = \arg\min_{\alpha} \sum_{i=1}^{n} L(y_i, f_{m-1}(x_i) + \alpha h(x_i;\ \theta_m)) \qquad (6)$$

where $L(\cdot)$ is the loss function measuring the difference between the prediction and the true value.

To prevent overfitting, a shrinkage parameter $\eta \in (0,1)$, also known as the learning rate, is introduced to scale the contribution of each tree:

$$f_m(x) = f_{m-1}(x) + \eta \cdot \alpha_m h(x;\ \theta_m) \qquad (7)$$

where a smaller value of η leads to slower learning but often yields better generalizability.

Part 3: Algorithmic details of Accumulated Local Effects (ALE)

Let *f(x)* denote the prediction function of the trained model, and let x_j be the feature of interests. The ALE method works by partitioning the domain of x_j into K intervals defined by grid points $z_0<z_1<\ldots<z_K$. For each interval $[z_{k-1}, z_k]$, ALE computes the average local effect as

$$\hat{f}_j'(z_k) = \frac{1}{n_k} \sum_{i \in I_k} \left[f\left(z_k, x_{i,\setminus j}\right) - f\left(z_{k-1}, x_{i,\setminus j}\right) \right] \qquad (8)$$

where $x_{i,\setminus j}$ is the feature vector for observation i excluding x_j, and I_k is the set of observations falling in the k-th bin (i.e., $x_{ij} \in [z_{k-1}, z_k]$).

The ALE up to point z_k is then given by

$$\mathrm{ALE}_j(z_k) = \sum_{l=1}^{k} \hat{f}_j'(z_l) \qquad (9)$$

This yields a function that describes how x_j contributes to model predictions across its range, averaged over the data distribution. The result is centered such that $\mathbb{E}\left[\mathrm{ALE}_j(x_j)\right] = 0$, ensuring interpretability in relative rather than absolute terms.

Part 4: Algorithmic details of Partial Dependence Plot (PDP)

Partial Dependence Plot (PDP) estimates the marginal effect of a target feature on the predicted outcome by averaging model predictions over the distribution of all other features.

Select two features x_j, x_k, we have the 2-D PDP measures as

$$PD_{jk}\left(x_j, x_k\right) = E_{x_{-(j,k)}}[\hat{f}(x_j, x_k, x_{-(j,k)})]$$

The result provides a global, model-agnostic view of feature-response relationships and is particularly useful for detecting non-linear patterns and thresholds.

Appendix B

Table B1. Model fit and comparisons.

Group	Model	Moran's I	AIC	RMSE	MAE	(Pseudo) R2
Overall	OLS	-	5195.4	21.54	15.24	0.439
	GWR	0.352***	5141.0	20.61	14.60	0.487
	GBDT	**-**	**-**	**14.14**	**9.42**	**0.800**
High	OLS	-	1290.2	21.57	16.43	0.476
	GWR	0.158*	1254.8	19.84	15.08	0.557
	GBDT	**-**	**-**	**18.81**	**13.09**	**0.657**
Middle	OLS	-	2669.4	22.45	15.96	0.399
	GWR	0.278***	2630.2	21.30	15.08	0.459
	GBDT	**-**	**-**	**15.80**	**10.51**	**0.759**
Low	OLS	-	1261.7	16.23	10.83	0.478
	GWR	0.132*	1222.0	14.46	9.88	0.585
	GBDT	**-**	**-**	**14.69**	**8.75**	**0.622**

Note: OLS refers to multiple linear regression solved by the ordinary least squares method; GWR denotes the geographically weighted regression model; and GBDT is the gradient boosting decision tree approach used in this study. Significance levels are indicated as *** $p<0.001$, ** $p<0.01$, and * $p<0.05$. For the GBDT model, the pseudo R^2 value represents the average performance across five-fold cross-validation.

Appendix C

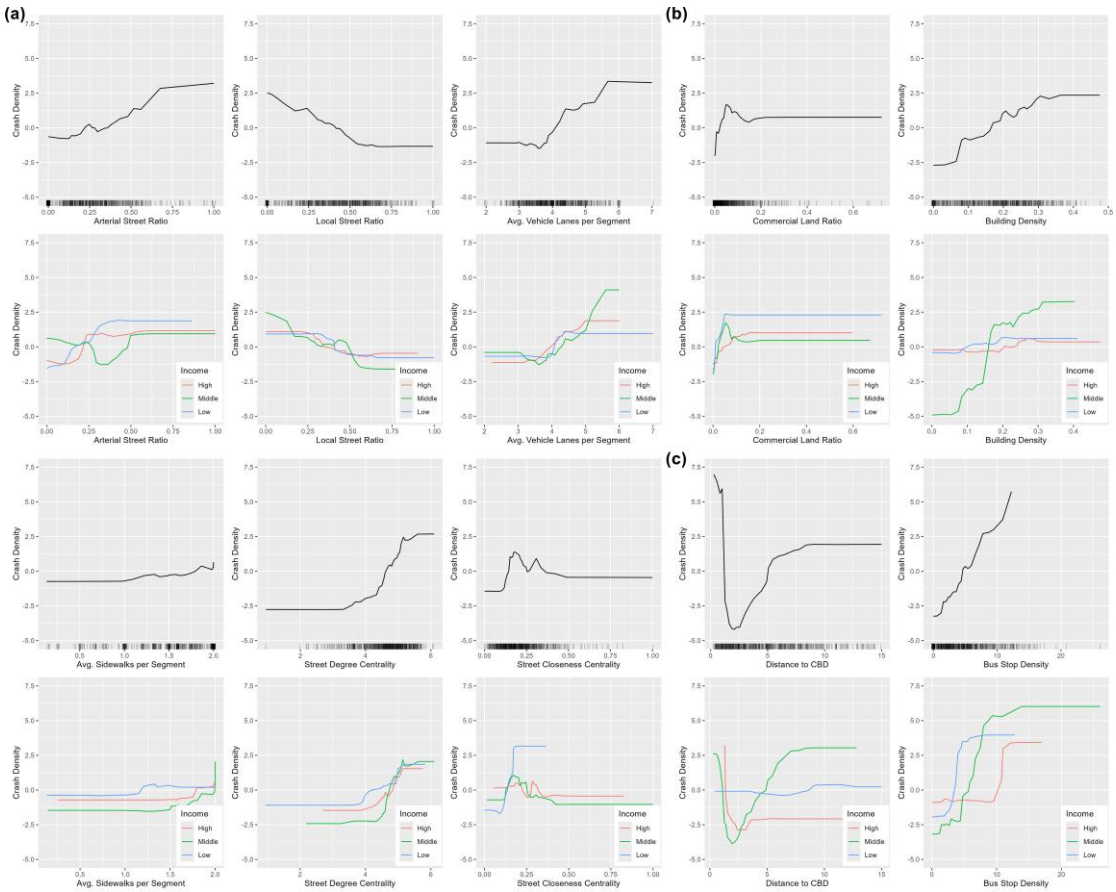


Figure C1. Effects of other variables on traffic crash density.

References

An, Rui, Zhaomin Tong, Yimei Ding, Bo Tan, Zihao Wu, Qiangqiang Xiong, and Yaolin Liu. 2022. "Examining non-linear built environment effects on injurious traffic collisions: A gradient boosting decision tree analysis." *Journal of Transport & Health* 24:101296. doi: 10.1016/j.jth.2021.101296.

An, Z. H., B. Xie, and Q. Y. Liu. 2023. "No street is an Island: Street network morphologies and traffic safety." *Transport Policy* 141:167-181. doi: 10.1016/j.tranpol.2023.07.023.

Barajas, Jesus M. 2018. "Not all crashes are created equal: Associations between the built environment and disparities in bicycle collisions." *Journal of Transport and Land Use* 11 (1). doi: 10.5198/jtlu.2018.1145.

Cai, Mingming, Junfeng Jiao, Minghai Luo, and Yanfang Liu. 2020. "Identifying transit deserts for low-income commuters in Wuhan Metropolitan Area, China." *Transportation Research Part D-Transport and Environment* 82. doi: 10.1016/j.trd.2020.102292.

Chakravarthy, B., C. L. Anderson, J. Ludlow, S. Lotfipour, and F. E. Vaca. 2010. "The Relationship of Pedestrian Injuries to Socioeconomic Characteristics in a Large Southern California County." *Traffic Injury Prevention* 11 (5):508-513. doi: 10.1080/15389588.2010.497546.

Chen, P. 2015. "Built environment factors in explaining the automobile-involved bicycle crash frequencies: A spatial statistic approach." *Safety Science* 79:336-343. doi: 10.1016/j.ssci.2015.06.016.

Chen, P., and Q. Shen. 2016. "Built environment effects on cyclist injury severity in automobile-involved bicycle crashes." *Accident Analysis and Prevention* 86:239-246. doi: 10.1016/j.aap.2015.11.002.

Chen, P., F. Y. Sun, Z. B. Wang, X. Gao, J. F. Jiao, and Z. M. Tao. 2018. "Built environment effects on bike crash frequency and risk in Beijing." *Journal of Safety Research* 64:135-143. doi: 10.1016/j.jsr.2017.12.008.

Cooper, Crispin H. V., and Alain J. F. Chiaradia. 2020. "sDNA: 3-d spatial network analysis for GIS, CAD, Command Line & Python." *SoftwareX* 12. doi: 10.1016/j.softx.2020.100525.

Ding, C., X. Y. Cao, M. X. Dong, Y. Zhang, and J. W. Yang. 2019. "Non-linear relationships between built environment characteristics and electric-bike ownership in Zhongshan, China." *Transportation Research Part D-Transport and Environment* 75:286-296. doi: 10.1016/j.trd.2019.09.005.

Ding, C., X. Y. Cao, and P. Naess. 2018a. "Applying gradient boosting decision trees to examine non-linear effects of the built environment on driving distance in Oslo." *Transportation Research Part A-Policy and Practice* 110:107-117. doi: 10.1016/j.tra.2018.02.009.

Ding, Chuan, Xinyu Cao, and Yunpeng Wang. 2018b. "Synergistic effects of the built environment and commuting programs on commute mode choice." *Transportation Research Part A-Policy and Practice* 118:104-118. doi: 10.1016/j.tra.2018.08.041.

Ding, Chuan, Peng Chen, and Junfeng Jiao. 2018c. "Non-linear effects of the built environment on automobile-involved pedestrian crash frequency: A machine learning approach." *Accident Analysis and Prevention* 112:116-126. doi: 10.1016/j.aap.2017.12.026.

Dommes, A., V. Cavallo, and J. Oxley. 2013. "Functional declines as predictors of risky street-crossing decisions in older pedestrians." *Accident Analysis and Prevention* 59:135-143. doi: 10.1016/j.aap.2013.05.017.

Dumbaugh, Eric, and Wenhao Li. 2011. "Designing for the Safety of Pedestrians, Cyclists, and

Motorists in Urban Environments." *Journal of the American Planning Association* 77 (1):69-88. doi: 10.1080/01944363.2011.536101.

Dumbaugh, Eric, and Yi Zhang. 2013. "The Relationship between Community Design and Crashes Involving Older Drivers and Pedestrians." *Journal of Planning Education and Research* 33 (1):83-95. doi: 10.1177/0739456x12468771.

Ewing, R., and E. Dumbaugh. 2009. "The Built Environment and Traffic Safety A Review of Empirical Evidence." *Journal of Planning Literature* 23 (4):347-367. doi: 10.1177/0885412209335553.

Feng, Z. X., N. Y. Ji, Y. Luo, N. N. Sze, J. F. Tian, and C. Zhao. 2021. "Exploring the influencing factors of public traffic safety awareness in China." *Cognition Technology and Work* 23 (4):731-742. doi: 10.1007/s10111-020-00655-8.

Friedman, J. H. 2001. "Greedy function approximation: A gradient boosting machine." *Annals of Statistic* 29 (5):1189-1232. doi: 10.1214/aos/1013203451.

Graham, D. J., and S. Glaister. 2003. "Spatial variation in road pedestrian casualties: The role of urban scale, density and land-use mix." *Urban Studies* 40 (8):1591-1607. doi: 10.1080/0042098032000094441.

Grise, Emily, Ron Buliung, Linda Rothman, and Andrew Howard. 2018. "A geography of child and elderly pedestrian injury in the City of Toronto, Canada." *Journal of Transport Geography* 66:321-329. doi: 10.1016/j.jtrangeo.2017.10.003.

Guerra, E., X. X. Dong, and M. Kondo. 2022. "Do Denser Neighborhoods Have Safer Streets? Population Density and Traffic Safety in the Philadelphia Region." *Journal of Planning Education and Research* 42 (4):654-667. doi: 10.1177/0739456x19845043.

Hu, S., H. F. Xing, W. Luo, L. Wu, Y. Y. Xu, W. M. Huang, W. K. Liu, and T. Q. Li. 2023. "Uncovering the association between traffic crashes and street-level built-environment features using street view images." *International Journal of Geographical Information Science* 37 (11):2367-2391. doi: 10.1080/13658816.2023.2254362.

Huang, Helai, Mohamed A. Abdel-Aty, and Ali Lotfi Darwiche. 2010. "County-level crash risk analysis in Florida: Bayesian spatial modeling." *Transportation Research Record* 2148 (1):27-37.

Jiang, Baoguo, Song Liang, Zhong-Ren Peng, Haozhe Cong, Morgan Levy, Qu Cheng, Tianbing Wang, and Justin V. Remais. 2017. "Transport and public health in China: the road to a healthy future." *Lancet* 390 (10104):1781-1791. doi: 10.1016/S0140-6736(17)31958-X.

Lee, S., J. Yoon, and A. Woo. 2020a. "Does elderly safety matter? Associations between built environments and pedestrian crashes in Seoul, Korea." *Accident Analysis and Prevention* 144. doi: 10.1016/j.aap.2020.105621.

Lee, Sugie, Junho Yoon, and Ayoung Woo. 2020b. "Does elderly safety matter? Associations between built environments and pedestrian crashes in Seoul, Korea." *Accident Analysis & Prevention* 144:105621. doi: 10.1016/j.aap.2020.105621.

Li, X., S. Y. Yu, X. Huang, B. Dadashova, W. C. Cui, and Z. Zhang. 2022. "Do underserved and socially vulnerable communities observe more crashes? A spatial examination of social vulnerability and crash risks in Texas." *Accident Analysis and Prevention* 173. doi: 10.1016/j.aap.2022.106721.

Ling, Changlong, Xinyi Niu, Jiawen Yang, Jiangping Zhou, and Tianren Yang. 2024. "Unravelling heterogeneity and dynamics of commuting efficiency: Industry-level insights into evolving

efficiency gaps based on a disaggregated excess-commuting framework." *Journal of Transport Geography* 115:103820. doi: 10.1016/j.jtrangeo.2024.103820.

Liu, F., Z. Zhang, H. Chen, and S. Nie. 2020. "Associations of ambient air pollutants with regional pulmonary tuberculosis incidence in the central Chinese province of Hubei: a Bayesian spatial-temporal analysis." *Environ Health* 19 (1):51. doi: 10.1186/s12940-020-00604-y.

Liu, Jinli, Gian Antariksa, Shriyank Somvanshi, and Subasish Das. 2025a. "Revealing equity gaps in pedestrian crash data through explainable artificial intelligence clustering." *Transportation Research Part D: Transport and Environment* 139:104538. doi: 10.1016/j.trd.2024.104538.

Liu, X. C., R. L. Qiao, Z. Q. Wu, T. R. Yang, X. N. Zhang, X. L. Zhang, and Z. L. Zhu. 2024. "Unveiling the spatially varied nonlinear effects of urban built environment on housing prices using an interpretable ensemble learning model." *Applied Geography* 173. doi: 10.1016/j.apgeog.2024.103458.

Liu, Y. X., J. J. Li, and W. Yang. 2025b. "Nonlinear interactive associations between built environments and adults' mental health in China using a longitudinal design." *Applied Geography* 180. doi: 10.1016/j.apgeog.2025.103662.

Luo, Yuanyuan, Yanfang Liu, Zhaomin Tong, Nannan Wang, and Lei Rao. 2023. "Capturing gender-age thresholds disparities in built environment factors affecting injurious traffic crashes." *Travel Behaviour and Society* 30:21-37. doi: 10.1016/j.tbs.2022.08.003.

Mansfield, Theodore J., Dana Peck, Daniel Morgan, Barbara McCann, and Paul Teicher. 2018. "The effects of roadway and built environment characteristics on pedestrian fatality risk: A national assessment at the neighborhood scale." *Accident Analysis and Prevention* 121:166-176. doi: 10.1016/j.aap.2018.06.018.

Marshall, S. 2004. *Streets and Patterns*. London, UK: Routledge.

Marshall, W. E., and N. W. Garrick. 2011. "Does street network design affect traffic safety?" *Accident Analysis and Prevention* 43 (3):769-781. doi: 10.1016/j.aap.2010.10.024.

Miranda-Moreno, L. F., P. Morency, and A. M. El-Geneidy. 2011. "The link between built environment, pedestrian activity and pedestrian-vehicle collision occurrence at signalized intersections." *Accident Analysis and Prevention* 43 (5):1624-1634. doi: 10.1016/j.aap.2011.02.005.

Morency, P., L. Gauvin, C. Plante, M. Fournier, and C. Morency. 2012. "Neighborhood Social Inequalities in Road Traffic Injuries: The Influence of Traffic Volume and Road Design." *Accident Analysis and Prevention* 102 (6):1112-1119. doi: 10.2105/AJPH.2011.300528.

Niebuhr, T., M. Junge, and E. Rosen. 2016. "Pedestrian injury risk and the effect of age." *Accident Analysis and Prevention* 86:121-128. doi: 10.1016/j.aap.2015.10.026.

Noland, R. B., N. J. Klein, and N. K. Tulach. 2013. "Do lower income areas have more pedestrian casualties?" *Accident Analysis and Prevention* 59:337-345. doi: 10.1016/j.aap.2013.06.009.

Noland, Robert B., and Maria Luz Laham. 2018. "Are low income and minority households more likely to die from traffic-related crashes?" *Accident Analysis and Prevention* 120:233-238. doi: 10.1016/j.aap.2018.07.033.

Pirdavani, Ali, Stijn Daniels, Karin van Vlierden, Kris Brijs, and Bruno Kochan. 2017. "Socioeconomic and sociodemographic inequalities and their association with road traffic injuries." *Journal of Transport & Health* 4:152-161. doi: 10.1016/j.jth.2016.12.001.

Qiao, S., A. G. O. Yeh, M. Z. Zhang, and X. Yan. 2020. "Effects of state-led suburbanization on traffic crash density in China: Evidence from the Chengdu City Proper." *Accident Analysis and Prevention* 148. doi: 10.1016/j.aap.2020.105775.

Quistberg, D. A., P. Hessel, D. A. Rodriguez, O. L. Sarmiento, U. Bilal, W. T. Caiaffa, J. J. Miranda, M. D. de Pina, A. Hernandez-Vasquez, and A. V. D. Roux. 2022. "Urban landscape and street-design factors associated with road-traffic mortality in Latin America between 2010 and 2016 (SALURBAL): an ecological study." *Lancet Planetary Health* 6 (2):E122-E131. doi: 10.1016/s2542-5196(21)00323-5.

Rifaat, S. M., R. Tay, and A. de Barros. 2010. "Effect of Street Pattern on Road Safety Are Policy Recommendations Sensitive to Aggregations of Crashes by Severity?" *Transportation Research Record* (2147):58-65. doi: 10.3141/2147-08.

Roll, J., and N. McNeil. 2022. "Race and income disparities in pedestrian injuries: Factors influencing pedestrian safety inequity." *Transportation Research Part D-Transport and Environment* 107. doi: 10.1016/j.trd.2022.103294.

Rong, Xiao, and Ying Jin. 2023. "How can large-scale housing provision in the outskirts benefit the urban poor in fast urbanising cities? The case of Beijing." *Cities* 132:104066. doi: 10.1016/j.cities.2022.104066.

Saha, Dibakar, and Eric Dumbaugh. 2021. "Use of a model-based gradient boosting framework to assess spatial and non-linear effects of variables on pedestrian crash frequency at macro-level." *Journal of Transportation Safety & Security*. doi: 10.1080/19439962.2021.1958036.

Saha, Dibakar, Eric Dumbaugh, and Louis A. Merlin. 2020. "A conceptual framework to understand the role of built environment on traffic safety." *Journal of Safety Research* 75:41-50. doi: https://doi.org/10.1016/j.jsr.2020.07.004.

Shao, Q. F., W. J. Zhang, X. Y. Cao, and J. W. Yang. 2022. "Nonlinear and interaction effects of land use and motorcycles/ E-bikes on car ownership." *Transportation Research Part D-Transport and Environment* 102. doi: 10.1016/j.trd.2021.103115.

Shin, Eun Jin. 2023. "Decomposing neighborhood disparities in bicycle crashes: A Gelbach decomposition analysis." *Transport Policy* 131:156-172. doi: 10.1016/j.tranpol.2022.12.014.

Soja, Edward W. 2013. *Seeking spatial justice*. Vol. 16: U of Minnesota Press.

Stoker, P., A. Garfinkel-Castro, M. Khayesi, W. Odero, M. N. Mwangi, M. Peden, and R. Ewing. 2015. "Pedestrian Safety and the Built Environment: A Review of the Risk Factors." *Journal of Planning Literature* 30 (4):377-392. doi: 10.1177/0885412215595438.

Tao, Tao, and Petter Næss. 2022. "Exploring nonlinear built environment effects on driving with a mixed-methods approach." *Transportation Research Part D-Transport and Environment* 111:103443. doi: 10.1016/j.trd.2022.103443.

Walker, Gordon. 2012. *Environmental justice: Concepts, evidence and politics*: Routledge.

Wang, J., and H. L. Huang. 2016. "Road network safety evaluation using Bayesian hierarchical joint model." *Accident Analysis and Prevention* 90:152-158. doi: 10.1016/j.aap.2016.02.018.

Wang, X. S., J. G. Yang, C. Lee, Z. R. Ji, and S. K. You. 2016. "Macro-level safety analysis of pedestrian crashes in Shanghai, China." *Accident Analysis and Prevention* 96:12-21. doi: 10.1016/j.aap.2016.07.028.

Wang, Xuesong, Jinghui Yuan, Grant G. Schultz, and Shouen Fang. 2018. "Investigating the safety

impact of roadway network features of suburban arterials in Shanghai." *Accident Analysis and Prevention* 113:137-148. doi: 10.1016/j.aap.2018.01.029.

WHO. 2023. Global Status Report on Road Safety 2023.

Wier, Megan, June Weintraub, Elizabeth H. Humphreys, Edmund Seto, and Rajiv Bhatia. 2009. "An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning." *Accident Analysis & Prevention* 41 (1):137-145.

Wilmut, K., and C. Purcell. 2022. "Why Are Older Adults More at Risk as Pedestrians? A Systematic Review." *Human Factors* 64 (8):1269-1291. doi: 10.1177/0018720821989511.

Wu, W. T., S. Y. Jiang, R. H. Liu, W. Z. Jin, and C. X. Ma. 2020. "Economic development, demographic characteristics, road network and traffic accidents in Zhongshan, China: gradient boosting decision tree model." *Transportmetrica A-Transport Science* 16 (3):359-387. doi: 10.1080/23249935.2020.1711543.

Xie, B., Z. H. An, Y. L. Zheng, and Z. G. Li. 2019. "Incorporating transportation safety into land use planning: Pre-assessment of land use conversion effects on severe crashes in urban China." *Applied Geography* 103:1-11. doi: 10.1016/j.apgeog.2018.12.003.

Xie, Bo, Changlong Ling, and Lan Wang. 2022. "Research on the impact of urban street pattern on traffic safety: a case study of Wuhan city." *City Planning Review*.

Xu, P. P., H. L. Huang, N. Dong, and M. Abdel-Aty. 2014. "Sensitivity analysis in the context of regional safety modeling: Identifying and assessing the modifiable areal unit problem." *Accident Analysis and Prevention* 70:110-120. doi: 10.1016/j.aap.2014.02.012.

Yang, J. W., J. Cao, and Y. F. Zhou. 2021. "Elaborating non-linear associations and synergies of subway access and land uses with urban vitality in Shenzhen." *Transportation Research Part A-Policy and Practice* 144:74-88. doi: 10.1016/j.tra.2020.11.014.

Yang, J. W., P. R. Su, and J. Cao. 2020. "On the importance of Shenzhen metro transit to land development and threshold effect." *Transport Policy* 99:1-11. doi: 10.1016/j.tranpol.2020.08.014.

Yu, C. Y., X. M. Zhu, and C. Lee. 2022. "Income and Racial Disparity and the Role of the Built Environment in Pedestrian Injuries." *Journal of Planning Education and Research* 42 (2):136-149. doi: 10.1177/0739456X18807759.

Yu, Chia-Yuan. 2014. "Environmental supports for walking/biking and traffic safety: Income and ethnicity disparities." *Preventive Medicine* 67:12-16. doi: 10.1016/j.ypmed.2014.06.028.

Yuan, Q., and J. Y. Wang. 2021. "Goods movement, road safety, and spatial inequity: Evaluating freight-related crashes in low-income or minority neighborhoods." *Journal of Transport Geography* 96. doi: 10.1016/j.jtrangeo.2021.103186.

Zhang, Yuanyuan, John Bigham, David Ragland, and Xiaohong Chen. 2015. "Investigating the associations between road network structure and non-motorist accidents." *Journal of Transport Geography* 42:34-47. doi: 10.1016/j.jtrangeo.2014.10.010.

Zheng, L., T. Sayed, and F. Mannering. 2021. "Modeling traffic conflicts for use in road safety analysis: A review of analytic methods and future directions." *Analytic Methods in Accident Research* 29. doi: 10.1016/j.amar.2020.100142.

Zhou, Wei, Pengpeng Xu, Jiabin Wu, and Junda Huang. 2025. "Examining macro-level traffic crashes considering nonlinear and spatiotemporal spillover effects." *Accident Analysis & Prevention* 211:107852. doi: 10.1016/j.aap.2024.107852.