



Multiscale geographical random forest: A novel spatial ML approach for traffic safety modeling integrating street-view semantic visual features

Pengfei Cui ^{a,b,c} , Mohamed Abdel-Aty ^c , Lei Han ^{c,*} , Xiaobao Yang ^b

^a College of Metropolitan Transportation, Beijing University of Technology, Beijing 100124, China

^b School of System Science, Beijing Jiaotong University, Beijing 100044, PR China

^c Department of Civil, Environmental & Construction Engineering, University of Central Florida, Orlando, FL 32816, United States

ARTICLE INFO

Keywords:

Traffic Safety
Spatial Machine Learning
Image Semantic Segmentation
Multiscale Geographical Random Forest
Visual Environment
Google Street View

ABSTRACT

Macro-level traffic safety modeling aims to identify critical risk factors to regional crashes, providing essential basis for effective countermeasures by traffic managers. Previous work mainly incorporated macro and static socio-demographic and infrastructure features, overlooking drivers' visual perception of environment, which crucially influences their driving behavior and thus safety. Moreover, spatial machine learning (ML) has gained prominence for its strong crash prediction performance. However, existing spatial ML typically apply spatial effects at a fixed or homogeneous scale (e.g., specific Euclidean distances), limiting their ability to capture the multiscale spatial heterogeneity of features. To address these gaps, emerging image semantic segmentation technique is employed to extract visual environment features (e.g., buildings, trees) from Google Street View (GSV) images. A novel spatial ML method, Multiscale Geographical Random Forest (MGRF), is proposed to overcome fixed-spatial scale constraints to adaptive multiscale spatial modeling. Empirical experiments on Southeast Florida show that the inclusion of visual environment features from 228,352 street view images leads to notably improved crash prediction. Compared to traditional models (e.g., multiscale geographically weighted regression), MGRF fits optimal spatial bandwidths for each sample, achieving improvements of 30.31%, 9.98%, and 5.53% in MSE, MAE, and R², respectively. By incorporating SHapley Additive ex-Planations, MGRF identified key risk features for each region and quantified their spatial heterogeneity. The Results reveal that in urban core areas, the proportion of cars in GSV, which reflects road traffic condition, is the most critical feature contributing positively to increase in crashes. In contrast, for suburban regions, lower road density and abundant green spaces are associated with a reduction in crashes. This study highlights the significant potential of integrating street-view semantic visual features with multiscale spatial ML to enhance traffic safety analysis.

1. Introduction

Traffic crashes represent a major public health concern, exerting substantial impacts on both health and economic stability

* Corresponding author.

E-mail addresses: pengfeicui@bjut.edu.cn (P. Cui), m.aty@ucf.edu (M. Abdel-Aty), le966091@ucf.edu (L. Han), yangxb@bjtu.edu.cn (X. Yang).

worldwide. The United Nations has proposed a goal to achieving a 50 % reduction in road traffic casualties by 2030 (WHO, 2022), underscoring the imperative to enhance traffic safety and protect community residents. However, in 2022, the United States recorded 42,795 traffic crash fatalities, with a fatal crash rate twice that of other countries (e.g., United Kingdom) (Department for Transport, 2023; NHTSA, 2023). In recent years, increasing research efforts have focused on macro-level modeling, specifically the analysis of aggregated crash frequency at the zonal level, to identify zone-level risk factors and devise effective countermeasures (Huang et al., 2010; Merlin et al., 2020; Tang et al., 2020; Yocum and Gayah, 2022). By establishing the relationship between regional crash frequency and potential contributing factors, these studies can identify critical factors and hotspots, thereby helping traffic managers implement targeted countermeasures and policies to prevent crashes (Wu et al., 2023a; Zheng et al., 2024; Wang et al., 2024b).

In existing studies, several kinds of features have identified as key influencing factors of regional traffic safety: socio-demographic features (Li et al., 2022), road network configuration (Huang et al., 2010; Merlin et al., 2020), and travel patterns (Lee et al., 2014). For instance, income level has been identified as a crucial socio-demographic feature. Higher-income regions typically have better road infrastructure and advanced safety features, contributing to lower fatal traffic crash rates (Wang et al., 2024; Yocum and Gayah, 2022). Other socio-economic indicators, including population density, poverty rate, and population age, have also been found to be strongly correlated with the crash rates (Azimian et al., 2021; Huang et al., 2010; Lee et al., 2023). Regarding road network configuration, factors such as higher road and intersection density, as well as segment-to-intersection ratios, have been found to be significantly associated with increased crashes and fatalities (Huang et al., 2010; Lee et al., 2015). Additionally, travel behaviors—including travel distance, transportation mode, and commute duration—are also influencing traffic crash rates (Rahman and Antipova, 2024). Regions with longer commuting distances tend to experience higher crash rates due to increased exposure to traffic hazards (Ewing et al., 2016).

Although existing features can capture overall traffic, demographic, and road geometric characteristics, they primarily represent the static infrastructure and macro-level traffic conditions, which ignore the micro drivers' visual perception of surrounding environment (Abdel-Aty et al., 2024; Cai et al., 2022). Leveraging advanced computer vision technique, recent studies directly extracted drivers' visual environment features from street images and videos, revealing the importance of the visual environments in influencing driver behaviors and safety (Abdel-Aty et al., 2024; Cai et al., 2020; Zhao and Khattak, 2018). For instance, Yu et al. (2019) and Abdel-Aty et al. (2024) have extracted semantic and distance features from the driver's dash camera videos and found that drivers are more likely to speed in an open field without trees and buildings. Similarly, Cai et al. (2022) proposed several visual measures from Google Street View (GSV) images to reflect drivers' visual environment and estimated its impact on speeding crash frequency on urban arterials. They concluded that the proportion of trees in drivers' view could reduce speeding crashes, but the complexity level of the visual environment could lead to more speeding crashes. These findings highlight the importance of incorporating the driver's visual environment's features into crash analysis to both improve model predictions and enhance interpretability (Cai et al., 2022; Fan et al., 2023a; Yu et al., 2019). However, there have been no studies exploring the impact of visual environment on macro-level traffic safety. Given that the visual environment (e.g., buildings, roads, roadside trees, etc.) vary significantly across regions, how to extract these complex and diverse regional visual environment features and investigate their impact on crashes still needs to be investigated.

As for the macro-level traffic safety modeling, spatial modeling has become essential for quantifying the spatial heterogeneity effects of features. Existing studies (Barua et al., 2016; Mannerling and Bhat, 2014; Wang et al., 2024) have shown that the impact of various factors on traffic crash frequency can vary significantly across regions. Traditional statistical methods, which assume independent observations and fixed covariate effects, fail to capture the complex interaction and spatial dependency inherent in traffic crash data. Therefore, spatial statistical methods such as random parameter regression model (Lee et al., 2023; Wang et al., 2025), spatial lag models (Hong et al., 2016), geographically weighted regression (GWR) (Li et al., 2022) and their variants have been widely applied to uncover complex spatial heterogeneity of covariates in traffic crash analysis. These spatial statistical methods, however, still rely on linear relationship assumption and struggle with increasing scale and complexity of traffic datasets (Wen et al., 2021; Zhou et al., 2023). Recent studies have shifted towards developing machine learning (ML) techniques, which have demonstrated significant effectiveness in modeling nonlinear and large-scale datasets (Fan et al., 2023). Approaches like support vector machine (Yu and Abdel-Aty, 2013), random forest (Wen et al., 2022), and gradient boosting (Cai et al., 2022) have shown better performance than statistical models.

However, existing ML models struggle to address the spatial heterogeneity in macro-level traffic crash modeling, as they treat all spatial data as 'independent' and 'aspatial' samples. Encouragingly, spatial ML methods have been proposed to integrate spatial heterogeneity effects into ML models. For example, the Geographically Random Forest (GRF) model, proposed by Georgano et al. (2021), establishes local random forests for each sample and its spatial neighbors to account for spatial heterogeneity. GRF has been utilized by Wu et al. (2024) and Wang et al. (2024) in traffic safety modeling and achieved notable prediction accuracy. Other spatial ML models such as Spatial Gaussian GPBoost, Graph Convolution Network have also been used in recent traffic studies (Trirat et al., 2023; Wu et al., 2023b; Han and Abdel-Aty, 2025). Nevertheless, existing spatial ML models mainly account for spatial effects at either a homogeneous or fixed spatial scale, making it struggling to handle the dynamic and multiscale spatial heterogeneity in real world traffic spatial data (Fotheringham et al., 2017). These simplifications limit the models' ability to accurately capture the dynamic nature of spatial influences in real-world settings. For example, crashes in urban areas may have stronger correlations with neighboring suburban traffic, while crashes in rural area are more influenced by local demographic factors, with less traffic interactions with nearby regions. Ignoring such heterogeneity in spatial scale can introduce biases for both model prediction and interpretation (Mannerling and Bhat, 2014; Tang et al., 2023). Therefore, it is essential to develop a novel spatial ML approach that can account for spatial heterogeneity following dynamic, real-world spatial scales in macro-level traffic crash frequency modeling.

With the above-mentioned research gaps, this study utilized advanced image segmentation technique to extract visual environment features and proposed a novel spatial ML approach, Multiscale Geographical Random Forest (MGRF), to capture their spatial

heterogeneity on regional crashes. Main contributions of this paper include:

- 1) Employing image semantic segmentation to extract visual environment features (e.g., building, sidewalk) from regional street views for macro-level traffic safety modeling.
- 2) Proposing MGRF to extend fixed-spatial scale constraints toward adaptive multiscale spatial modeling, enabling more precise spatial heterogeneity representation to enhance both model prediction and interpretability.
- 3) Incorporating explainable SHapley Additive exPlanations (SHAP) framework with MGRF to identify key features and their contributions to regional crash frequency.

The paper is organized into five sections. Following this section, [section 2](#) presents the data preparation. [Section 3](#) shows the details of the proposed methodology and [section 4](#) illustrates the experiment results. Finally, the conclusions are presented in [Section 5](#).

2. Data preparation

2.1. Research area and data sources

This study focuses on Southeast Florida, specifically targeting Miami-Dade, Broward, and Palm Beach counties, which are among the areas in Florida with the highest traffic crashes frequency as shown in [Fig. 1 \(a\)](#). The research area, highlighted in red in [Fig. 1 \(b\)](#), includes the urban and suburban zones where most residents live, deliberately excluding the western, less inhabited vegetative areas.

Three types of datasets were utilized in this study: (1) Traffic crash records spanning 2021–2023 were obtained from the Signal Four Analytics system of Florida Department of Transportation, offering a three-year window to capture recent traffic safety trend. (2) For socio-demographic data, we incorporated the Social Vulnerability Index (SVI) data from the U.S. Department of Health and Human Services. The SVI provides a quantitative measure of relative social vulnerability across census tracts, enabling the integration of demographic and economic variables into our spatial analysis. (3) To extract the visual environment features along regional roadways, we first obtained road network GIS data for the study area from OpenStreetMap. Then, we systematically sampled points along these road segments and captured street view images for each location using Google Street View. Advance image semantic segmentation technique was used to extract the visual environment features (e.g., road, sky, car, buildings, terrain, etc.). The details of each data processing are elaborated in the following sections.

2.2. Spatial region Creation and crash aggregation

Prior to conducting macro-level spatial modeling, it is necessary to divide the study area of Southeast Florida into basic study regions. Numerous scholars have examined the factors influencing regional traffic safety while retaining traditional geographic unit (such as census tract ([Li et al., 2022](#)), zip code ([Lee et al., 2014](#)), traffic analysis zones ([Siddiqui et al., 2012](#))). However, such spatial units typically use road sections as boundaries, leading to many traffic crashes being indexed ambiguously when they occur on region

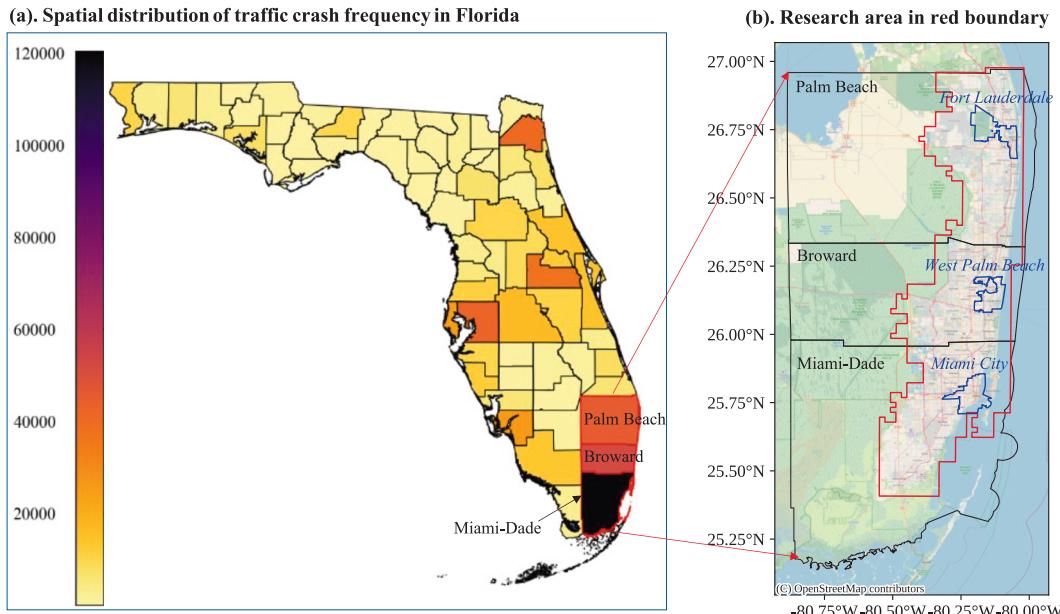


Fig. 1. Research area in Southeast Florida (Miami-Dade, Broward, and Palm Beach county).

boundaries. In recent year, grid-level segmentation, which divides space into equal-sized cells, has been widely employed in traffic safety and demonstrated superior stability and scalability in standardization (Cui et al., 2024; Fan et al., 2023b; Ren et al., 2018). Thus, we divided the study area into equal-sized square grids using the grid-based partitioning method. We evaluated several grid sizes, ranging from $1\text{ km} \times 1\text{ km}$ to $5\text{ km} \times 5\text{ km}$, as illustrated in Fig. 2.

As the grid size increased, a more pronounced aggregation of traffic crashes became evident, reflecting broader distribution trends. To identify the optimal spatial resolution for spatial partitioning, we evaluated five grid scales ($1\text{ km} \times 1\text{ km}$ to $5\text{ km} \times 5\text{ km}$) using a composite metric integrating three spatial statistics. First, we quantified the strength of spatial autocorrelation via Moran's I (Bivand et al., 2009):

$$\text{Global Moran's I} = \frac{N}{W} \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (1)$$

where N is the total number of grid cells, y_i denotes the observed crash frequency in grid i , and \bar{y} is the mean crash frequency across all grids. The spatial weight w_{ij} equals 1 if cells i and j share a common boundary or vertex, and 0 otherwise, such that $W = \sum_{i=1}^N \sum_{j=1}^N w_{ij}$. Higher values of I indicates stronger clustering of crashes.

Second, we computed the hotspot overlap ratio $H(s, s')$, following (Harris et al., 2017), which measures the spatial consistency between hotspot regions at two scales:

$$H(s, s') = \frac{|\mathcal{H}_s \cap \mathcal{H}_{s'}|}{|\mathcal{H}_s|} \quad (2)$$

Where \mathcal{H}_s comprises all grid cells g_i at scale s whose crash density meets or exceeds the 90th percentile of densities at that scale (that is, the top 10 % by density). The numerator $|\mathcal{H}_s \cap \mathcal{H}_{s'}|$ counts the number of cells identified as hotspots at both scales s and s' . This ratio captures how well the most intense crash-density areas align when the grid resolution changes.

Third, we assessed intra-grid heterogeneity using local variance withing 3×3 neighborhoods (Drăguț et al., 2011):

$$V = \frac{1}{N} \sum_{i=1}^N \left[\frac{1}{9} \sum_{j \in \mathcal{N}(g_i)} \left(\text{Density}(g_i) - \frac{1}{9} \sum_{j \in \mathcal{N}(g_i)} \text{Density}(g_j) \right)^2 \right] \quad (3)$$

where $\mathcal{N}(g_i)$ denotes the set of nine grids forming the 3×3 neighborhood centered on grid g_i , and $\text{Density}(g_i)$ is the crash density in grid g_i at scale s .

Finally, the three statistics were combined into a single composite score for each scale s as:

$$\text{Score}(s) = \underbrace{\left(\frac{I_s}{\max(I)} \times 0.4 \right)}_{\text{Autocorrelation}} + \underbrace{\left(\frac{H_s}{\max(H)} \times 0.4 \right)}_{\text{Hotspot overlap ratio}} + \underbrace{\left(\frac{\min(V)}{V_s} \times 0.2 \right)}_{\text{Local variance}} \quad (4)$$

where I_s , H_s , and V_s are the values of Moran's I, hotspot overlap ratio, and local variance at scale s with weights of 0.4, 0.4, and 0.2 assigned to emphasize cluster detection over variance minimization. By ranking each scale according to this composite score, we

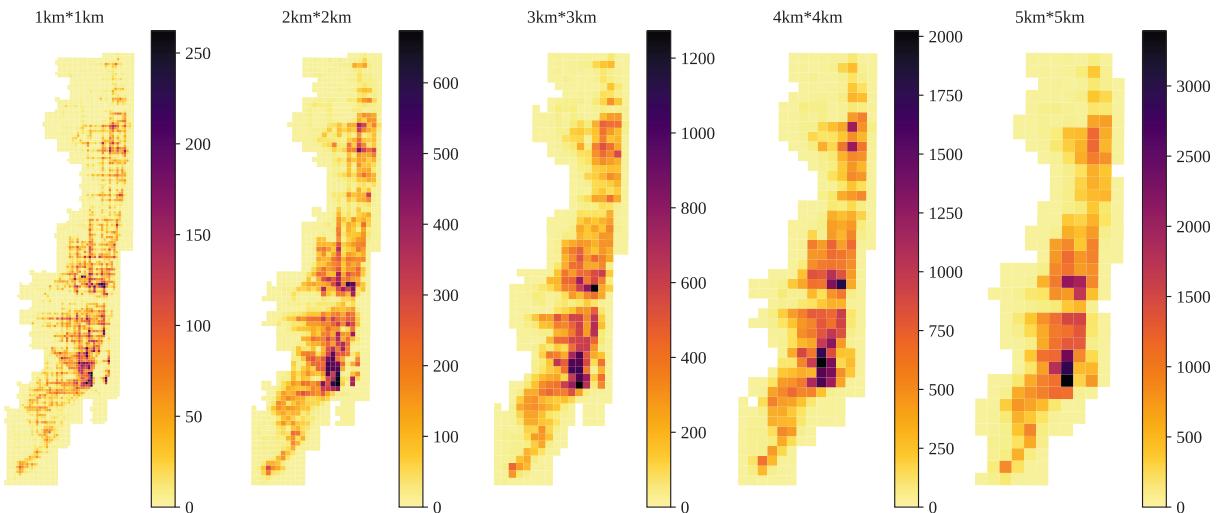


Fig. 2. Spatial distribution of crash frequency at different gird level.

determine the grid size that best balances clustering strength, hotspot stability and within-cell homogeneity. Table 1 summarizes the performance metrics across the different spatial resolutions.

The 2 km resolution achieved the highest composite score (0.96), driven by its peak spatial autocorrelation ($I = 0.142$, $p < 0.001$) and near-maximal hotspot overlap ratio (0.97). While the 1 km grid exhibited comparable clustering strength ($I = 0.134$, $p < 0.001$), its larger sample size ($n = 5383$) imposed higher computational costs without improving discriminative power. In contrast, coarser grids (≥ 3 km) exhibited progressive degradation; for example, Moran's I became non-significant at 4 km ($I = 0.033$, $p = 0.21$), and hotspot ratios declined by 55 %–58 % at 4–5 km.

Ultimately, we determined that a 2 km \times 2 km grid offered an optimal balance for spatial partitioning. This resolution effectively preserves overarching crash distribution patterns across Southeast Florida while retaining sufficient detail to reveal localized heterogeneity. The selected grid size thus facilitates capturing essential variations without excessive aggregation, ensuring that both global scale trends and local scale spatial differences are robustly represented in the analysis.

2.3. Socio-demographic features spatial matching

Given the SVI data were originally collected at the census tract level, a transformation was required to align it with the grid-based analytical framework used in this study. To address this, we implemented a matching procedure that assigns population and SVI values to each individual grid cell (Cui et al., 2025), as illustrated in Fig. 3. The overlaying process allowed for the derivation of grid-level features from the original census tract data as follows:

$$PP_i = \sum_{j=1}^n \frac{S_{ij}}{S_j} \times PP_j \quad (5)$$

$$SVI_i = \frac{\sum_{j=1}^n \left(SVI_j \times \frac{S_{ij}}{S_j} \times PP_j \right)}{\sum_{j=1}^n \left(\frac{S_{ij}}{S_j} \times PP_j \right)} \quad (6)$$

where S_{ij} denotes the area of overlap between census tract j and grid unit i , S_j and PP_j is the total area and population size in census tract j , SVI_j is the original social vulnerability for census tract j . For each grid unit i , the weighted SVI index was calculated by aggregating the SVI values from all intersecting census tracts, weighted by their overlapping areas and population values. This grid-based transformation process, as depicted in Fig. 3, effectively alleviating the disparities between densely populated and sparsely populated areas, ensuring that socio-demographic characteristics are represented uniformly across all grid cells. Our weighted grid approach also provides a more nuanced spatial representation, capturing socio-economic variations with greater precision, which ultimately strengthens the analysis by mitigating biases that might arise from traditional census tract-based aggregation.

2.4. Visual environment features extraction

As shown in the Fig. 4, the visual environment features were extracted from Google Street View (GSV) images along the roads. First, the road network GIS files were obtained through OpenStreetMap and matched with each split research grid. The street view points were then generated every 100 m along the roadways using ArcMap 10.6, totaling 57,088 points. For each street view point, four GSV images at 0° (north), 90° (west), 180° (south), and 270° (east) directions were collected and spliced as a full GSV image to comprehensively coverage the road environment. To make sure the GSV images similar to the drivers' view, the horizontal field of view and pitch were set as 50° and 0°, respectively (Cai et al., 2022; Li et al., 2015; Yang et al., 2009). Finally, a total of $4 \times 57,088 = 228,352$ GSV images were collected for the whole road network. To extract the visual environment features, Segmenter, a transformer-based segmentation model, was used to cluster objects from images. Proposed by Strudel et al. (2021), Segmenter integrates global context information into image segmentation, outperforming convolution-based methods to achieve state-of-the-art. Here, a total of 18 types of objects in the drivers' view environment were labeled in different colors (e.g., road, sidewalk, sky, building, fence, vegetation, terrain, and vehicle).

Based on the semantic segmentation results, the object type by each pixel could be determined in each GSV image. Referring to existing studies (Abdel-Aty et al., 2024; Cai et al., 2022; Han et al., 2025), the proportion of each object type $Prop_k$ was calculated:

$$Prop_k = \frac{P_k}{\sum_{k=1}^{18} P_k}, k = 1 : 18 \quad (7)$$

Where P_k is the total pixel number of object type k , which can be sky, road, buildings, vegetation, vehicle, or sidewalk. After obtaining the object proportion of each GSV image, the data is aggregated at the research grid level. To capture both the average and variation level of these visual features across the entire grid, the mean, median, 25th percentile, 75th percentile, and standard deviation of each object proportion are calculated. For example, Car_median_i represents the proportion of car in the GSV images of the i th research grid. Additionally, the number of GSV images in each grid is used as an exposure variable to represent the total road length, which is closely related to crash frequency. Finally, a total of $18 \times 5 + 1 = 91$ visual environment features are extracted as potential modeling features.

Table 1

Performance metrics across spatial resolutions.

Spatial grid	Sample size	Moran's I (p-value)	Hotspot overlap ratio (vs. 1 km)	Local variance	Composite score
1 km × 1 km	5383	0.134***	1.00	29.88	0.93
2 km × 2 km	1420	0.142***	0.97	26.37	0.96
3 km × 3 km	675	0.101 **	0.81	23.22	0.76
4 km × 4 km	417	0.033	0.58	30.55	0.52
5 km × 5 km	287	0.056*	0.55	23.11	0.53

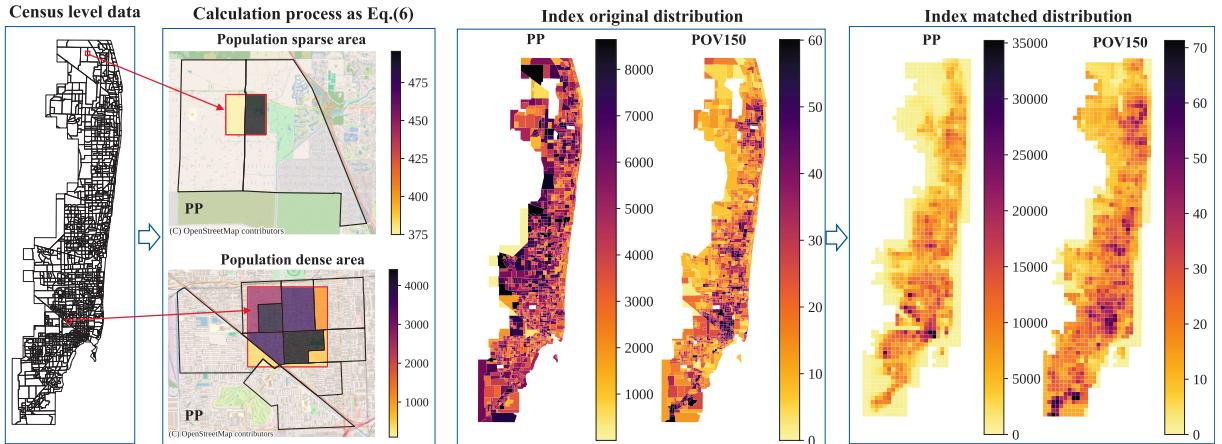


Fig. 3. Process of socio-demographic features spatial matching.

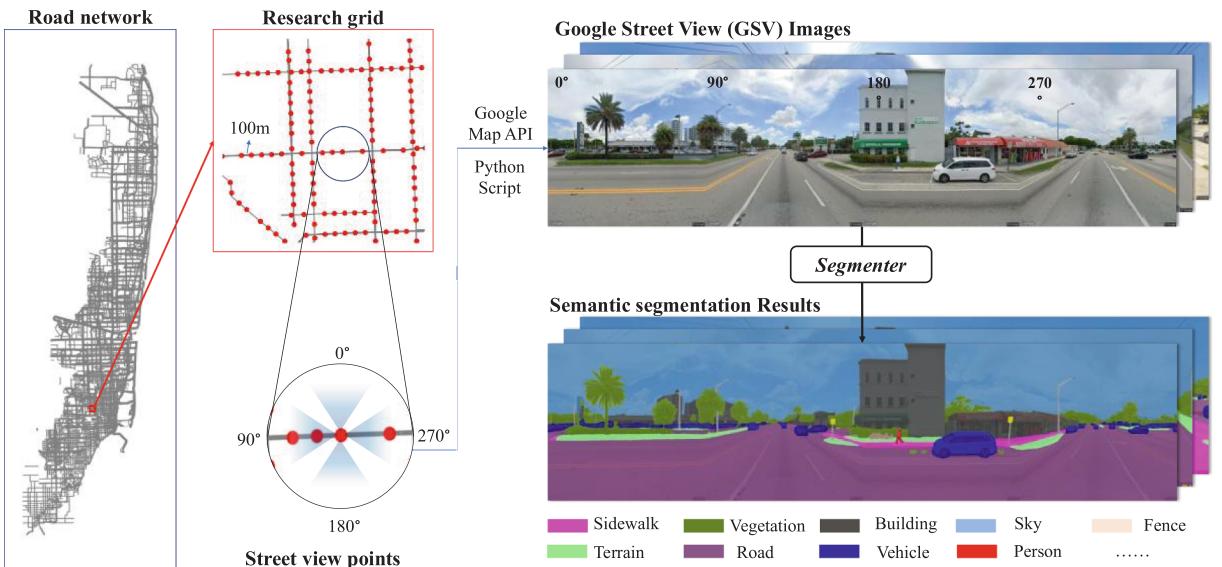


Fig. 4. The framework of visual environment features extraction from GSV images.

2.5. Important features selection

To exclude unrelated features and increase model efficiency, feature selection is a key step to identify the most relevant features. Recently, random forest has been widely used as a powerful tool for feature selection (Speiser et al., 2019; Sun et al., 2023; Yu et al., 2021). Therefore, a random forest model was established to first select the top-20 important features for regional crashes. Then, a Pearson's correlation test was conducted among these 20 features, excluding highly correlated features (correlation coefficient >

± 0.65) to avoid multicollinearity (Mukaka, 2012; Naznin et al., 2016). It is noted that we also integrated traffic features such as annual average daily traffic (AADT), road length from FDOT GIS. However, we found that AADT data were incomplete at nearly 10 % of research areas. Additionally, regional AADT and road length are significantly proportional to the number of GSV images in study grid (Sample_count) (correlation coefficient are 0.65 and 0.84). Therefore, we opted to use the Sample_count as a regional exposure feature to effectively reflect both regional road length and AADT without introducing additional data sources. This approach ensured data consistency while enhancing the model's applicability and scalability across diverse urban contexts. Ultimately, a total of 11 features, 5 socio-demographic features and 6 visual environment features, were selected as the modeling features, with their statistical summaries provided in Table 2.

3. Methodology

3.1. Multiscale Geographical random forest (MGRF)

The proposed MGRF framework, illustrated in Fig. 5, consists of three key steps: (1) model fitting using multiscale bandwidths, (2) spatial-weighted ensemble prediction, and (3) model interpretation with SHAP. In the first step, both global and local models are fitted based on training samples to establish the general and spatial-specific correlations between features and targets. The second step uses these trained models, weighted by spatial proximity, to make predictions on unseen test samples and enhance prediction performance. While the last model interpretation combines the feature importance from local RF with SHAP-based feature correlations to quantify the spatial heterogeneity of each feature. The details of each step are elaborated further in the following sections. Moreover, we have made our complete implementation of the MGRF framework openly available on GitHub,¹ along with comprehensive documentation and a step-by-step example to facilitate adoption and further development by the research community. Compared to existing methods, the MGRF offers three key advantages:

- (1) The use of multiscale bandwidths allows each local RF to fit the optimal neighbors and capture heterogeneous spatial patterns in a flexible and effective manner.
- (2) By ensembling multiple nearby local RFs, the spatially weighted local prediction approach reduces prediction bias due to one single local RF model.
- (3) The integration of RF with SHAP enables the identification of both feature importance (i.e., degree of impact) and correlations (i.e., positive or negative effects).

3.1.1. Model fitting with multiscale bandwidths

In traditional RF algorithm, it treats all data as 'independent' and 'aspatial' samples to hardly address spatial heterogeneity (Deng et al., 2020; Georganos et al., 2021; Gu et al., 2023). It can be written as a simplistic version of a regression equation:

$$Y_i = a\mathbf{X}_i + \epsilon, i = 1 : n \quad (8)$$

where Y_i is the value of the dependent variable for the i th observation; $a\mathbf{X}_i$ is the nonlinear prediction of RF based on a set of \mathbf{X}_i independent variable vectors; ϵ denotes an error term. It is clear that the above equation is formed by using all the data at the same time, disregarding their spatial distribution.

Inspired by geographically weighted regression (Fotheringham et al., 1997), Georganos et al (2021) proposed the GRF to overcome such limitation. In GRF, for each observation i , a local RF is computed only including several of its spatially nearby observations. Thus, the Equation (8) is extended to:

$$Y_i = a(u_i, v_i)\mathbf{X}_i + \epsilon, i = 1 : n \quad (9)$$

where $a(u_i, v_i)\mathbf{X}_i$ is the prediction of a local RF model calibrated on location i , and (u_i, v_i) are the corresponding coordinates. The area that the local RF operates in is called the 'neighborhood' (or 'kernel'), and the maximum distance from a data point to its kernel is known as the 'bandwidth' (Brunsdon et al., 1998). Two common types of kernels are used to define neighborhoods: 'adaptive' and 'fixed'. In the former, the neighborhood is defined by the n nearest neighbors and in the latter, by a circle whose radius is the bandwidth (Brunsdon et al., 1998; Georganos et al., 2021). As shown in the Fig. 6, for the sample \mathbf{X}_i , the bandwidth d_i determines the neighborhood boundary. Only sample within this neighborhood can be selected (green points) to fit the local RF of \mathbf{X}_i .

However, the GRF algorithm assumes a fixed bandwidth d_i for all data points, limiting each sample to operate at a uniform spatial scale. This assumption neglects the potential heterogeneity and variability inherent in the spatial-specific locations and data scales (Fotheringham et al., 2017; Gu et al., 2023). For example, urban areas may exhibit strong traffic correlations with the nearby suburban areas, thus requiring a larger bandwidth to extend their spatial scales for traffic safety model fitting. In contrast, a rural area might require a smaller bandwidth since crashes there are mainly influenced by local conditions. Therefore, referring to the Multiscale Geographically Weighted Regression (Fotheringham et al., 2017), we relax the fixed-bandwidth assumption by allowing different processes to operate at varying spatial scales, enabling model fitting with multiscale bandwidths:

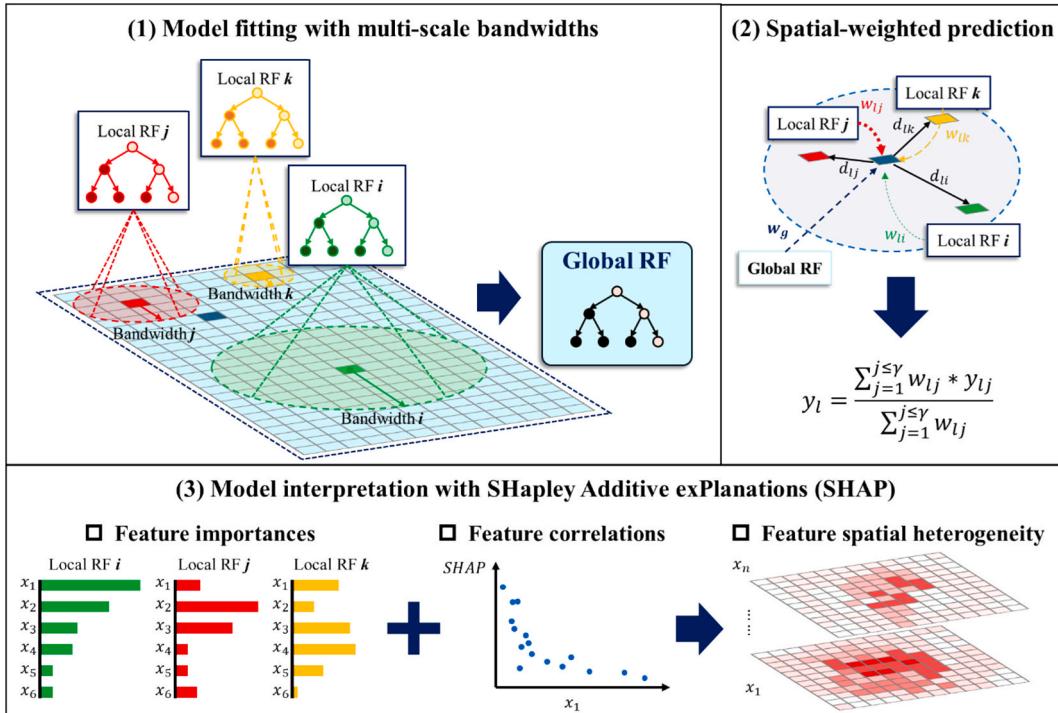
$$Y_i = a(u_i, v_i, d_i)\mathbf{X}_i + \epsilon, i = 1 : n \quad (10)$$

¹ The Python code is publicly available at https://github.com/PengfeiCui99/Multiscale_Geographical_Random_Forest_MGRF.

Table 2

Statistical summarizes of modeling features.

Features	Description	Min	Max	Mean	Std
<i>Socioeconomic features</i>					
PP	Population size in study grid	3.99	3.52e4	5909.83	4649.81
EP_NOVEH	Percentage of households with no vehicle available (%)	0	43.04	5.53	5.01
EP_POV150	Percentage of person below 150 % poverty (%)	1.50	70.56	18.88	10.63
EP_CROWD	Percentage of occupied housing units with more people than rooms (%)	0	23.57	4.33	3.61
EP_LIMENG	Percentage of person (age 5 +) who speak English "less than well" (%)	0	51.32	9.89	9.23
<i>Visual environment features</i>					
Sample_count	Number of GSV images in study grid	1	239	53.87	31.27
Car_median	The median car proportion in GSV images (%)	0	14.10	0.55	0.87
Person_q25	The 25 % percentile of person proportion in GSV images (%)	0	0.12	3.02e-3	6.43e-3
Sidewalk_q75	The 75 % percentile of sidewalk proportion in GSV images (%)	0	11.52	2.51	1.76
Fence_q25	The 25 % percentile of fence proportion in GSV images (%)	0	3.62	0.11	0.23
Terrain_q25	The 25 % percentile of terrain (grass/ greenery) proportion in GSV images (%)	0.11	23.77	6.72	4.91

**Fig. 5.** The framework of the MGRF model.

where $a(u_i, v_i, d_i)$ now reflects the non-linear function of local RF model, incorporating the spatial coordinates and bandwidth d_i .

The key step in the proposed MGRF is to select the optimal bandwidths for each local RF. To achieve that, an Out-Of-Bag (OOB)-based local bandwidth optimization algorithm is developed as shown in the **Pseudo Algorithm 1**. Without prior knowledge on the best bandwidth distribution, the core idea of this algorithm is to iteratively adjust the bandwidth value to explore the optimal one that yields the best local fitting performance:

- (1) Firstly, the upper and lower bound of the bandwidth d_{up} , d_{low} and the step size Δd are defined as hyperparameters. For each training sample X_i in Ω_{train} , its best MSE_i , bandwidth d_i , and Local RF can be initialized as ∞ , d_{low} , and NULL, respectively.
- (2) Then, a temporary bandwidth d_t is first set to d_{low} , and the algorithm iteratively adjusts its value until it exceeds the upper bound d_{up} . In each iteration, a set of neighbors Ω_t defined by the current bandwidth d_t is constructed, and a random forest RF_t is trained. The OOB mean squared error MSE_t is then calculated. If MSE_t is lower than the current best MSE_i , the algorithm will update MSE_i , RF_i , and the optimal bandwidth d_i . Meanwhile, the bandwidth d_t is incremented by Δd to enter the next iteration.
- (3) After iterating through all bandwidth values, the optimal random forest model RF_i and the corresponding bandwidth d_i are stored in RF_{train} and d_{train} , respectively. This process is repeated for all samples in Ω_{train} , resulting in a set of local random forest models with their corresponding multiscale bandwidths.

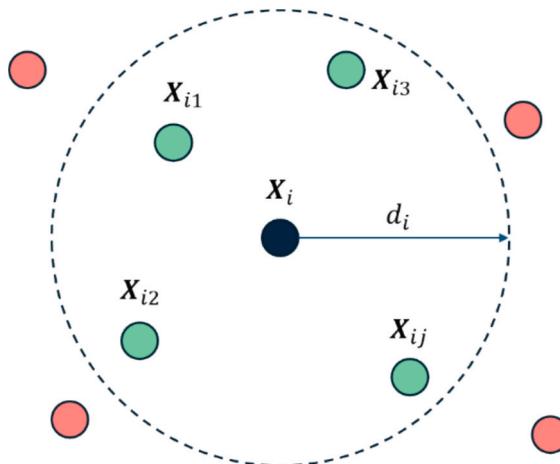


Fig. 6. The illustration of neighborhood of sample X_i with bandwidth.. d_i

Pseudo Algorithm 1: Local Bandwidth Optimization

Input:

d_{up} , d_{low} : the upper and lower bound of bandwidth; Δd : the step size of increasing bandwidth; X_i : the i th train sample; Ω_{train} : a set of n training samples

1: Initialize $RF_{train} \leftarrow []_n$, $d_{train} \leftarrow []_n$

2: **for** each X_i in Ω_{train} **do**

3: If Criterion = MSE, set

4: Initialize $MSE_i = \infty$, $d_i = d_t = d_{low}$, $RF_i = NULL$

4: **while** $d_t \leq d_{up}$ **do**

5: $\Omega_t = (X_{i1}, \dots, X_{ij}) \leftarrow (X_i, d_t)$ // select neighbors given bandwidth of d_t

6: $RF_t \leftarrow$ train random forest given Ω_t

7: $MSE_t \leftarrow$ OOB MSE // out-of-bag mean square error from RF_t

8: if $MSE_t < MSE_i$, do

9: $MSE_i \leftarrow MSE_t$, $RF_i \leftarrow RF_t$, $d_i \leftarrow d_t$

10: $d_t \leftarrow d_t + \Delta d$

11: $RF_{train}.append(RF_i)$, $d_{train}.append(d_i)$

13: **end for**

Output:

RF_{train} : a set of n local RFs with multiscale bandwidths; d_{train} : a list of n optimal bandwidths

Importantly, while our pseudocode defaults to using out-of-bag (OOB) mean squared error (MSE) as the optimization criterion, we have also implemented alternative metrics that enable optimization based on R2 or Moran's I. In the case of Moran's I, the algorithm first computes the residuals—defined as the difference between observed and predicted values—and then constructs a spatial weight matrix using a k-nearest neighbors approach via the KDTree algorithm (Hou et al., 2018). This spatial weight matrix is subsequently used to calculate Moran's I for the residuals, along with the corresponding p-value and z-score, thereby providing a quantitative assessment of spatial autocorrelation. By offering the option to select among MSE, R2, and Moran's I, our framework allows users to tailor the optimization process according to their analytical priorities, whether that be maximizing predictive accuracy or explicitly incorporating spatial correlation.

In addition to the local RF models with varying bandwidths, a global RF is also established using the entire training samples Ω_{train} . This model captures the overall relationships between features and targets that may not be significantly influenced by spatial locations. The combination of global and local RFs not only improves prediction accuracy but also enhances the interpretability of both micro-level and macro-level spatial patterns (Fotheringham et al., 2017; Georganos et al., 2021).

3.1.2. Spatial-weighted ensemble prediction

In traditional GRF (Georganos et al., 2021), it combines the prediction from the local RF models that is closest to the test data and the prediction from the global RF model using a weighted approach:

$$\hat{Y}_l = w_g \hat{Y}_{gi} + (1 - w_g) \hat{Y}_{li}, 0 < w_g < 1 \quad (11)$$

where \hat{Y}_l is the final prediction for the i th observation; \hat{Y}_{gi} and \hat{Y}_{li} are the predictions of global RF and closest local RF model, respectively. w_g represents a weight between 0 and 1 to balance the weights between these two models.

However, the \hat{Y}_{li} relies solely on the closest local RF model, making it highly sensitive to the quality of that particular model. For instance, if an outlier exists in the local region, it could significantly impact the trained local RF model. Additionally, other nearby RF

models (e.g., the second and third closest) could provide valuable information for prediction as well, yet they are not utilized in the current GRF model. Based on these considerations, we propose a spatially weighted ensemble approach for local prediction. As shown in Fig. 7(a), we first set the pre-defined upper bandwidth (d_{up}) for X_l to find a set of candidate trained local RFs (e.g., RF_i , RF_j , RF_k) within this bandwidth. Then, the distances between X_l and each candidate RFs (i.e., d_i , d_j , d_k) are compared against the respective optimal bandwidths of the trained RFs (i.e., d_i , d_j , d_k). Only those RFs for which the distance (d_{*l}) is less than or equal to their corresponding optimal bandwidth (d_*) are selected as the final prediction RF sets (e.g., RF_i , RF_k). This can make sure:

(1) The prediction of out-of-sample can identify all potential relevant local RFs ($d_{jl} \leq d_{up}$) to capture more useful localized information.

(2) Local RFs that lack significant spatial correlation ($d_{jl} \leq d_j$) are excluded, preventing erroneous associations between the test sample and unrelated local models.

Finally, the ensemble of their predictions in a spatially weighted manner using Equation (12):

$$\widehat{Y}_{li} = \frac{\sum_{j \in \Omega_l} w_{jl} * \widehat{Y}_{jl}}{\sum_{j \in \Omega_l} w_{jl}} = \frac{\sum_{j \in \Omega_l} w_{jl} * RF_j(\mathbf{X}_l)}{\sum_{j \in \Omega_l} w_{jl}}, \quad (12)$$

$$\text{where } \Omega_l := \{j | d_{jl} \leq d_{up} \text{ and } d_{jl} \leq d_j\}$$

where \widehat{Y}_{jl} is the prediction from the j th nearby local RF model RF_j given the features \mathbf{X}_l of the j th out-of-sample; w_{jl} is the spatial weight determined by the distance between locations l and j . Following existing studies (Deng et al., 2020; Fotheringham et al., 2017; Sun et al., 2024), we use the widely used ‘bisquare’ function/kernel to compute w_{jl} :

$$w_{jl} = \left(1 - \left(\frac{d_{jl}}{d_{up}}\right)^2\right)^2 \quad (13)$$

As shown in Fig. 7(b), the bisquare function assigns weights based on the distances and decreasing the weight smoothly as the distance increases. Therefore, the predictions from closer local RF models are assigned higher weights than those from farther away local RF models. Since this spatially weighted local prediction combines the predictions from all local RF models within the bandwidth, it is more robust and less susceptible to data outliers affecting one particular local RF model.

3.1.3. Model interpretation with SHapley additive exPlanations (SHAP)

Given that the MGRF combines a global RF and several local RFs, it can provide feature importances at both global and local levels (Breiman, 2001; Georganos et al., 2021; Gu et al., 2023). Specifically, RF model uses two criteria to measure the quality of a classification tree: the ‘Gini index’ and ‘Entropy-based information gain’. These indicators reflect how each feature contributes to an average decrease in error in the model classification process. The average of these criteria from all trees serves as a measure of feature importance. However, feature importance ranking alone is insufficient for explaining individual predictions. The contribution of each feature to the target is still unknown (Han et al., 2024b; Kim and Kim, 2022). In other words, while RF can quantify the degree of impact that features have on the target, it cannot determine whether the influence is positive or negative, which is equally important for interpretation.

Therefore, we introduce the widely used SHAP framework to both global and local RFs to quantify how each feature contributes to the target. The SHAP, proposed by Lundberg and Lee (2017), aims to describe the performance of a machine learning model based on game theory (Štrumbelj and Kononenko, 2014) and local explanations (Ribeiro et al., 2016). It offers an easy and effective measure to estimate the feature contributions and has been widely utilized in machine learning interpretation studies (Cui et al., 2025; Han et al.,

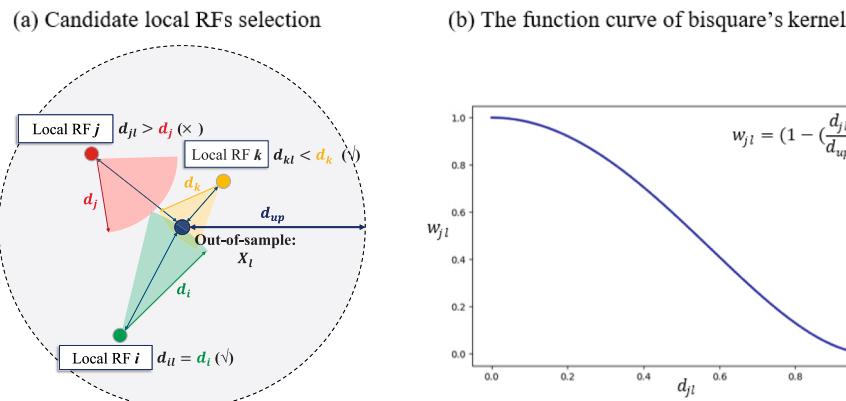


Fig. 7. Spatial-weighted ensemble prediction.

2024a; Yu et al., 2024). Assume a RF model where a group F (with n features) is used to predict an output. In SHAP, the contribution of each feature on the model output $f(F)$ is allocated based on its marginal contribution (Lundberg and Lee, 2017). The SHAP value ϕ_i of the i th feature is calculated through:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} S! \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)] \quad (14)$$

where F represents the set of all features and S represents all feature subsets obtained from F after removing the i th feature. $\frac{|S|!(|F| - |S| - 1)!}{|F|!}$ represents the probability weight of S calculated after feature permutation and combination. $f_{S \cup \{i\}}$ and f_S represent the i th feature and the prediction value of the model after S input, and predictions of these two models are compared to the current input $f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)$, where x_S represents the values of the input features in the set S . To estimate ϕ_i , the SHAP approach approximates the Shapley value by either performing Shapley sampling or Shapley quantitative influence.

In this study, the SHAP analysis is conducted to explore interactions between the features and targets at both global and local RFs. The SHAP interpretations are performed using TreeExplainer (Lundberg et al., 2020).

3.2. Baselines and model evaluation metrics

To evaluate the performance of our proposed MGRF, we compared it against six widely used baseline models in traffic safety research: Geographically Weighted Regression (GWR), Multiscale Geographically Weighted Regression (MGWR), Spatial BART, Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and Multilayer Perceptron (MLP). GWR is a classical spatial regression model for analyzing spatially varying relationships (O'Sullivan, 2003; Tang et al., 2020), while MGWR extends this approach by optimizing local bandwidths (Fotheringham et al., 2017). Spatial Bayesian additive regression trees (BART), which incorporates Bayesian additive regression trees with spatial autoregression, offers a flexible method for capturing multiscale heterogeneity and maintaining interpretability (Krueger et al., 2020). In addition, RF and XGBoost are robust tree-based machine learning models for handling nonlinearities, and MLP is a neural network that learns feature weights iteratively through backpropagation. For these ML models, their hyperparameters are turned using random grid search to get the optimal values as shown in Table 3.

To evaluate the prediction performance of candidate models, three measures include the Mean Absolute Error (MAE), Mean Squared Error (MSE), and R^2 . For, the lower the values of MAE and MSE and the higher the value of R^2 indicate the better model in terms of prediction performance. These three measures can be calculated as follow:

$$MAE = \frac{1}{N} \sum |Y_i - \hat{Y}_i| \quad (15)$$

$$MSE = \frac{1}{N} \sum (Y_i - \hat{Y}_i)^2 \quad (16)$$

$$R^2 = 1 - \frac{\sum (Y_i - \hat{Y}_i)^2}{\sum (Y_i - \bar{Y}_i)^2} \quad (17)$$

Table 3
The optimal model hyperparameters setting.

Models	Hyperparameters	Tuning range	Selected value
RF	Number of estimators	20, 50, 100, 200	100
	Maximum tree depth	10, 20, 30	20
	Maximum features	'sqrt', 'log2', '1/3'	'1/3'
XGboost	Objective function	'squarederror', 'squaredlogerror'	'squarederror'
	Learning rate	0.1, 0.001, 0.0001	0.1
	Number of estimators	5, 10, 20, 30	5
	Maximum tree depth	10, 20, 30	20
	Layer number	1, 2, 4, 6	2
	Nodes of 1st, 2nd layer	16, 32, 64, 128	32 & 16
MLP	Loss function	MSE, MAE, HuberLoss	MSE
	Learning rate	5e-3, 1e-3, 5e-4, 1e-4	1e-4
	Batch size	32, 64, 128, 256	64
	Number of trees	50, 100, 200, 400	100
	Prior scaling parameter	1, 2, 3	3
	Degrees of freedom	1, 3, 5	3
Spatial BART	Number of MCMC chains	1, 2, 4, 8	4
	Burn-in iterations	200, 500, 1000	500

4. Results

4.1. Model performance comparison

To assess model performance, the dataset was randomly divided into training (70 %) and testing (30 %) subsets, with all models trained on the former and evaluated on the unseen test set. Table 4 presents a detailed comparison of spatial prediction baselines and MGRF on the test set, evaluated with three key metrics: MSE, MAE, and R^2 . Among these baselines, the XGBoost model performs the weakest, with the highest MSE (2963.43), MAE (35.16), and the lowest R^2 (0.7502), indicating that it may be inadequate for capturing complex spatial dependencies. While accounting for potential spatial heterogeneity to some extent (Li et al., 2022), it still lagged behind more flexible nonlinear models, such as MLP and RF. The MGWR model improves upon traditional GWR by allowing adaptive bandwidth optimization, thereby enabling a more refined spatial adaptation (Fotheringham et al., 2017). Nonetheless, its parametric nature may restrict the full capture of nonlinear relationships in crash data. Spatial BART, a tree-based spatial regression method, further enhanced predictive performance by flexibly modeling spatially varying relationships without imposing strict parametric assumptions (Krueger et al., 2020). It outperformed MGWR and other baselines, demonstrating its ability to effectively balance spatial adaptivity and model complexity. However, Spatial BART does not explicitly optimize bandwidth selection, which may limit its ability to fully exploit localized spatial structures.

Overall, both the fixed-bandwidth GRF and the proposed MGRF achieved even greater accuracy improvements, reinforcing the effectiveness of integrating spatial heterogeneity with ML techniques (Georganos et al., 2021; Gu et al., 2023; Wang et al., 2024). The proposed MGRF model outperforms all other models, achieving the lowest MSE (1592.79), lowest MAE (27.08), and highest R^2 (0.8462). Compared to the baselines, it demonstrates significant improvements of 30.31 %, 9.98 %, and 5.83 % in MSE, MAE, and R^2 , respectively.

Compared to the fixed-bandwidth GRF, MGRF allows the modeling bandwidth to adaptively vary across spatial locations, resulting in further improvements of 9.67 % in MSE, 3.03 % in MAE, and 1.23 % in R^2 . To demonstrate the benefits of multiscale bandwidth in spatial modeling, ablation experiments were conducted to compare the fixed-bandwidth GRF and multiscale-bandwidth MGRF. For the MGRF model, the lower and upper bound of the bandwidth d_{low} , d_{up} were set to 30, and 140, respectively, with a step size Δd of 10. Based on the proposed local bandwidth optimization algorithm, each local RF in MGRF can search for the optimal bandwidth within the range 30–140. In contrary, 12 separate GRF models with fixed bandwidth values from 30, 40, to 140 were also trained to reflect the model performance with fix bandwidths. Fig. 8 presents their performance evaluation in terms of MSE and the R^2 across varying bandwidths. The predictive performance of MGRF model is depicted by the red dashed line. The performance evaluation, in terms of MSE and R^2 , shows that the MGRF model demonstrates superior predictive capability compared to all fixed bandwidth GRF models. The adaptive bandwidth strategy employed by MGRF consistently yields lower MSE and higher R^2 values near optimal bandwidth settings compared to the fixed bandwidth method, underscoring the advantages of adapting to spatial heterogeneity.

To validate the effectiveness of MGRF across diverse spatial contexts, prediction performance was evaluated across datasets spanning spatial scales from 1 km × 1 km to 5 km × 5 km, as detailed in Table 5. While MGRF's improvements over GRF in certain metrics—such as mean absolute error (MAE) at 2 km × 2 km (27.08 vs. 27.62)—might appear modest, the model demonstrates consistent superiority across all scales in critical indicators: mean squared error (MSE) is reduced by 13.5 %, 7.4 %, 14.7 %, 19.2 %, and 11.2 % at 1 km × 1 km, 2 km × 2 km, 3 km × 3 km, 4 km × 4 km, and 5 km × 5 km, respectively, reflecting enhanced calibration to scale-specific variance. Concurrently, R^2 values exhibit uniform increases, with the most substantial gains observed at 1 km × 1 km (0.6766 to 0.7203) and 4 km × 4 km (0.8407 to 0.8714), indicating stronger explanatory power for scale-dependent relationships. These trends underscore MGRF's capacity to capture nuanced spatial patterns across granularities, mitigating systematic error and advancing the characterization of scale-sensitive associations—key contributions that solidify the methodological rigor of the proposed multiscale framework.

We also compared the training times of the proposed models and baseline methods as Fig. 9 shown to evaluate the balance between computational efficiency and predictive performance. The results reveal marked disparities in computational complexity between spatial modeling approaches and conventional machine-learning methods. For instance, tree-based models such as XGBoost and RF require minimal training time (approximately 6.4 s for XGBoost and 4.4 s for RF), whereas spatial models like GWR and MGWR

Table 4
Model performance comparison between baselines and MGRF.

Models		MSE	MAE	R^2
Baselines	XGBoost	2963.43	35.16	0.7502
	GWR	2807.21	34.83	0.7634
	MLP	2745.21	31.01	0.7686
	RF	2739.58	30.33	0.7691
	MGWR	2321.64	29.78	0.7823
	Spatial BART	2075.58	29.54	0.7996
Fixed-bandwidth GRF		1720.52	27.62	0.8339
		(−20.64 %)*	(−6.95 %)*	(+4.29 %)*
Multiscale GRF (MGRF)		1592.79	27.08	0.8462
		(−30.31%)*	(−9.98 %)*	(+5.83 %)*

*: Improvement relative to the optimal performance of the baselines.

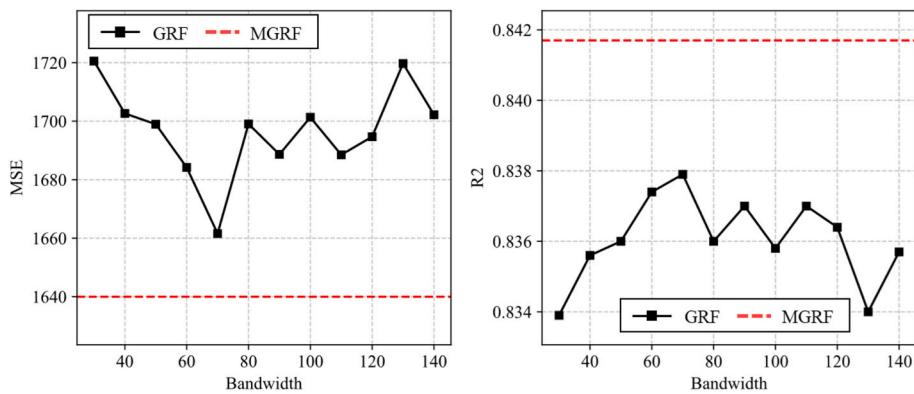


Fig. 8. Comparison of MSE and R^2 for Multiscale and Fixed bandwidth models.

Table 5

Model performance comparison across datasets between GRF and MGRF.

Datasets	GRF			MGRF		
	MSE	MAE	R^2	MSE	MAE	R^2
1 km × 1 km	298.69	8.50	0.6766	258.38	8.04	0.7203
2 km × 2 km	1720.52	27.62	0.8339	1592.79	27.08	0.8462
3 km × 3 km	7002.47	47.15	0.8227	5974.81	43.48	0.8487
4 km × 4 km	11435.84	58.31	0.8407	9237.14	52.99	0.8714
5 km × 5 km	33510.43	100.88	0.8250	29743.28	89.03	0.8447

demand substantially longer durations—with MGWR taking around 486.6 s. Spatial BART also exhibits higher computational costs (96.0 s) than simpler ML methods but remains more efficient than GRF and MGRF. Although the GRF and MGRF approaches also incur higher computational costs (approximately 362.4 and 341.2 s, respectively), these increases are offset by significant improvements in predictive performance. Both GRF and MGRF more effectively capture spatial heterogeneity than fixed-scale models, thereby providing superior local predictions. Notably, whereas conventional GRF necessitates an iterative search over fixed bandwidth values, MGRF employs an adaptive bandwidth selection process that simultaneously evaluates multiple spatial scales to determine the optimal bandwidth for each local model. This strategy enables MGRF to more precisely capture local spatial heterogeneity, thereby enhancing predictive accuracy without incurring further computational overhead relative to GRF.

Furthermore, Fig. 10 compares the local R^2 between the GRF with a fixed bandwidth of 70 and MGRF model, showing a marked

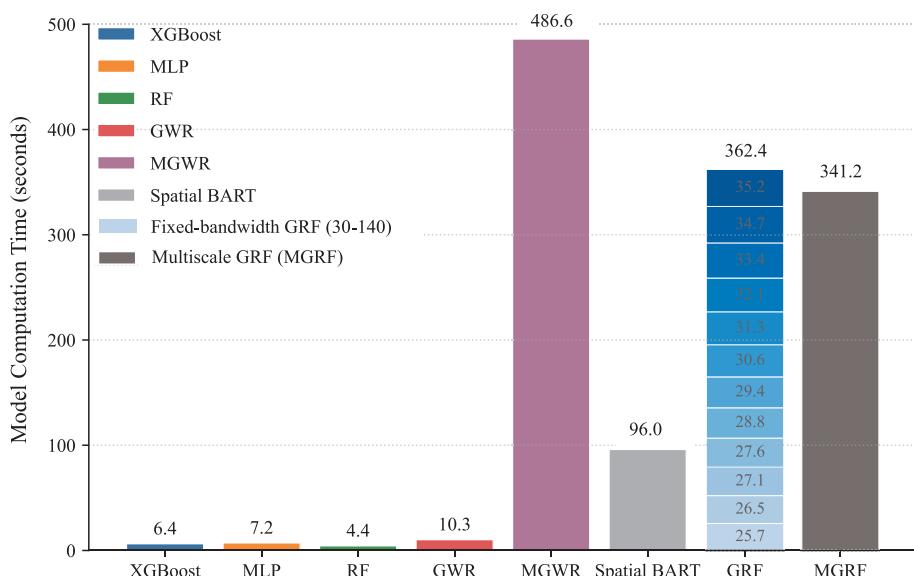


Fig. 9. Computational complexity of different models.

improvement in model fit, as consistently higher R^2 values (redder points) are observed throughout the spatial domain. The right panel, which shows the adaptively selected bandwidth values ranging from 30 to 140, underscores the ability of MGRF to dynamically adjust bandwidths to reflect local spatial heterogeneity. This adaptability allows the MGRF model to achieve a superior fit by tailoring bandwidths to varying local conditions, ultimately resulting in more accurate and refined spatial predictions compared to the fixed bandwidth approach. The adaptive bandwidth strategy offers a significant advantage by better capturing complex and heterogeneous spatial relationships, enhancing both local and global model accuracy. Overall, such results highlight the efficacy of multiscale bandwidths in capturing heterogeneous and feature spatial patterns, thereby improving the accuracy and robustness of spatial predictions.

Another ablation experiment was done to investigate the benefits of introducing visual environment features into traffic safety analysis. The results in [Table 6](#) show that, compared to using only social-demographic features, the model with visual environment features alone achieves lower MSE (34.79), lower MAE (31.78), and higher R^2 (0.7882). It indicates that visual environment features derived from GSVs can directly reflect drivers' visual perception, thus may show greater impact on driving behaviors and safety. Overall, the model with both types of features performed best (MSE = 1592.79, MAE = 27.08, R^2 = 0.8462). On the one hand, it demonstrates that beyond traditional social-economic features, the introduction of visual environment features offers new insights to identify potential risk factors (e.g., human perception, driving environment) related to traffic safety; On the other hand, it emphasizes the importance of integrating multiple data sources, particularly visual information, to enhance model accuracy in spatial prediction tasks.

4.2. Feature spatial heterogeneity interpretation

In this section, we delve into the complex spatial dynamics of features influencing crash frequency based on the results from MGRF model. To thoroughly investigate the spatial heterogeneity of visual environment and socioeconomic features, we focus on both feature importance (reflecting the degree of impact) and feature contributions (indicating positive or negative effects). The former is directly obtained from local RFs in the MGRF, while the later are derived from SHAP results of MGRF. Specifically, we first examine the spatial variation in feature importance, identifying key features and their regional significance ([Section 4.2.1](#)). Building on this, we then employ SHAP values and their spatial distributions to reveal the positive or negative effects of each feature on crash frequency, providing a comprehensive understanding of their impact across different geographic contexts ([Section 4.2.2](#)).

4.2.1. Feature importance with spatial variation

To ensure the robustness and interpretability of our proposed MGRF model, a comparative analysis of feature importances was conducted against high-performing baselines Spatial BART. As shown in [Fig. 11](#), both models consistently identified key predictors, including sample_count (importance scores: Spatial BART = 0.17, MGRF = 0.14), PP (Spatial BART = 0.10, MGRF = 0.12), and Car_median (Spatial BART = 0.09, MGRF = 0.15), confirming their fundamental relevance to the analysis. However, Spatial BART

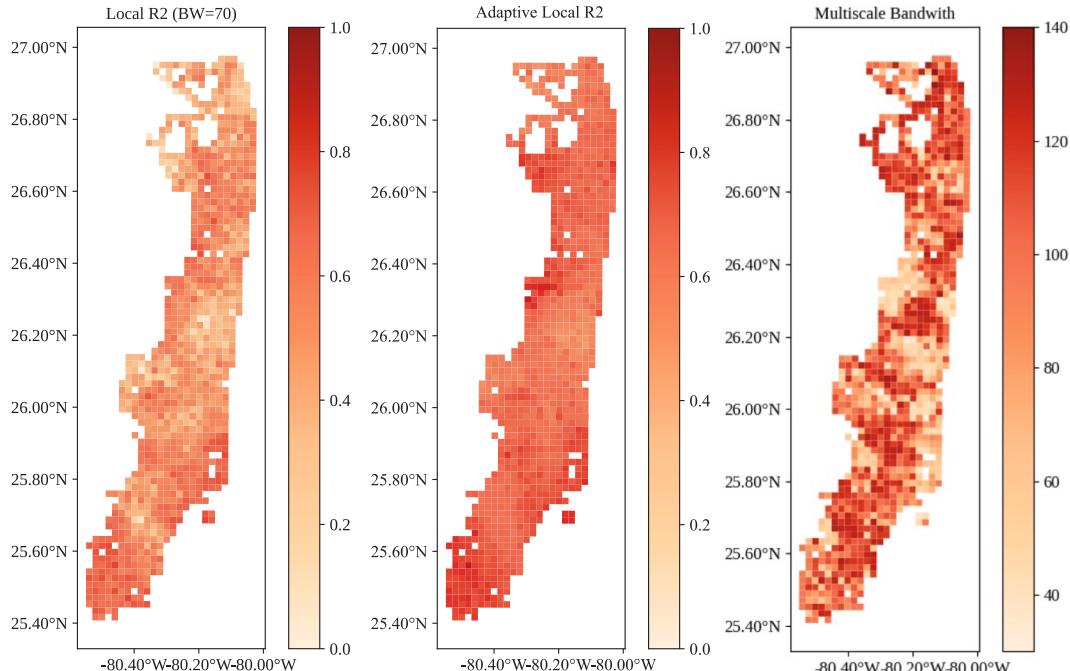


Fig. 10. Local R^2 comparison and multiscale bandwidths in MGRF.

Table 6

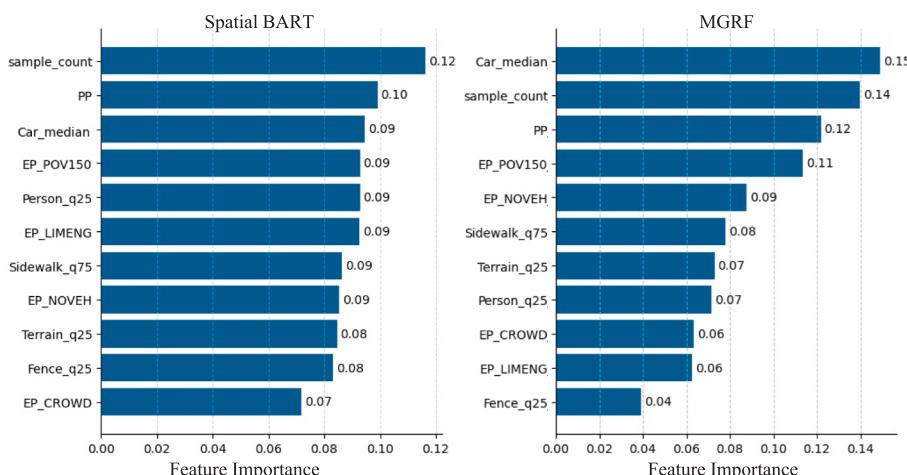
Ablation experiment of social-demographic and visual environment features on MGRF performance.

Features	MSE	MAE	R ²
Socio-demographic features (S)	2740.48	34.79	0.7354
Visual environment features (V)	2194.00	31.78	0.7882
With S + V	1592.79	27.08	0.8462

exhibited limited discriminative capacity, as multiple covariates (e.g., Car_median, EP_POV150, Person_q25) were confined within a narrow importance range (0.08–0.09), precluding meaningful hierarchical interpretation of their contributions. In contrast, MGRF enabled clearer differentiation of feature hierarchies, allowing unambiguous ranking of predictor significance. Notably, MGRF emphasized the visual-environmental feature (Car_median) as the dominant regional-scale factor. The enhanced predictive accuracy of MGRF, combined with its stable extraction of feature relationships, justifies its adoption for subsequent analyses. These results validate MGRF's utility in balancing interpretability and predictive performance within complex modeling contexts.

Referring to existing studies (Georganos et al., 2021; Gu et al., 2023; Wu et al., 2024), the Mean Gini index of features from each local RF in MGRF is utilized to represent their local importances to crash frequency among the 1042 research grids in the Southeast Florida. Fig. 12 visualizes the spatial variability in socio-demographic and visual environment features' importance for predicting regional crash frequency, showing that the importances of these features differ across each grid due to their different spatial pattern and traffic conditions. Boxplots in Fig. 12 (a) illustrates the distribution of local importance values for the 11 modeling features. Overall, Car_median, sample count (number of GSV images in study grid), and population size exhibited the highest median importance values, underscoring their crucial role as key exposure factors influencing crash frequency. The proportion of households without vehicles (EP_NOVEH) and the proportion of persons below 150 % poverty (EP_POV150) also demonstrate significant contributions, underscoring the role of social vulnerability in predicting regional crash frequency (Lin et al., 2021; Odijk et al., 2023). Fig. 12 (b) maps the spatial distribution of the most important feature for each region, revealing the geographic heterogeneity of factors driving crashes. Statistical analysis indicates that Car median and sample count are the most critical features in approximately 27.35 % and 19.77 % of regions, respectively, particularly in densely populated urban core areas. This finding aligns with the established correlation between vehicle ownership, road activity levels, and increased crash risk. Total population (PP) emerged as the most important feature in 13.92 % of regions, typically in highly populated areas. Socio-demographic features, specifically the proportion of persons below 150 % poverty (EP_POV150) and the proportion of households without vehicles (EP_NOVEH), were identified as the most important features in 9.69 % and 9.21 % of regions, respectively. These factors were primarily concentrated in Miami city and the Fort Lauderdale area, suggesting that residents in economically vulnerable communities face elevated traffic safety risks. In contrast, environment factors such as Terrain 25th percentile (Terrain_q25) and Sidewalk coverage 75th percentile (Sidewalk_q75) exhibited greater influence in peripheral or less densely populated regions, serving as the dominant features in approximately 8.83 % and 7.87 % of regions, respectively. The statistical insights from Fig. 12 (b) underscore the significant spatial heterogeneity in feature importance, emphasizing the necessity for tailored interventions that address the unique risk factors present in each area. The observed spatial heterogeneity in feature importance highlights the need for geographically tailored interventions in traffic safety modeling and policy development.

To further present the spatial heterogeneity of modeling features, Fig. 13 visualizes the spatial distribution of these feature importances across all regions in Southeast Florida. The results reveal significant spatial variations in the importance of different features for predicting crashes across geographic locations. Notably, certain features, such as Car median, consistently demonstrate high predictive importance across most regions. Socio-demographic factors, including poverty rates and the proportion of households

**Fig. 11.** Comparison of global feature importance in model Spatial BART and MGRF.

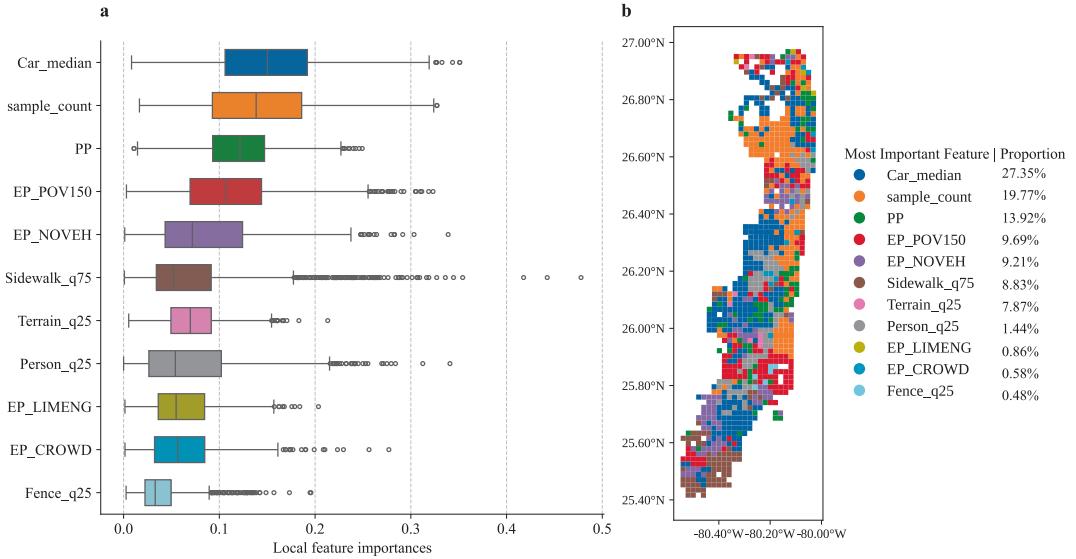


Fig. 12. Local feature importance ranking and spatial distribution of most important features.

without vehicles, exhibit particularly high predictive power in specific areas. When overlaid with the original feature distributions, these high-importance regions often coincide with urban, densely populated areas characterized by higher levels of socioeconomic vulnerability.

4.2.2. Shap-based feature contributions

To complement our feature importance analysis and provide a more nuanced understanding of feature contributions, we employed the SHAP framework to interpret the positive or negative impact of each feature on crash frequency. Fig. 14 presents a comprehensive visualization of feature impacts obtained through SHAP analysis. The scatter plots provide an intuitive representation of each feature's contributions, with SHAP values distributed along the horizontal axis and point colors transitioning from blue to red to indicate increasing feature values.

Among the visual environment features, median car proportion in GSVs (Car_median), 25 % percentile of person proportion in GSVs (Person_q25), 75 % percentile of sidewalk proportion in GSVs (Sidewalk_q75), all exhibits clear positive influences on crash frequency. Specifically, increases in these features generally corresponded to higher SHAP values, suggesting that they contribute to

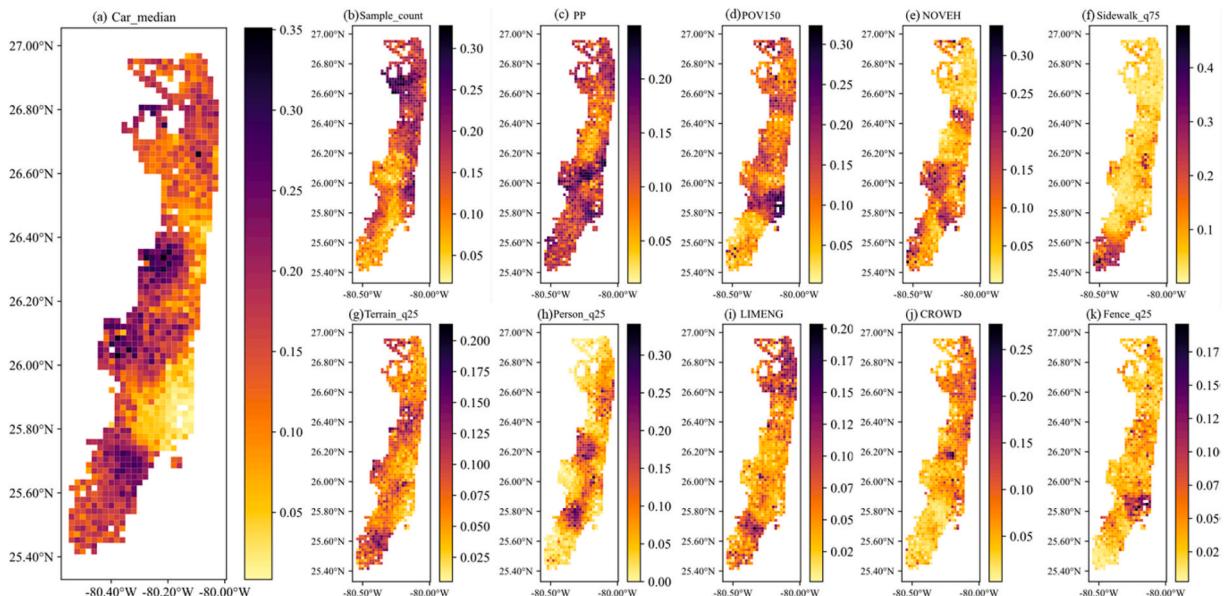


Fig. 13. Spatial distribution of feature importances in Southeast Florida.

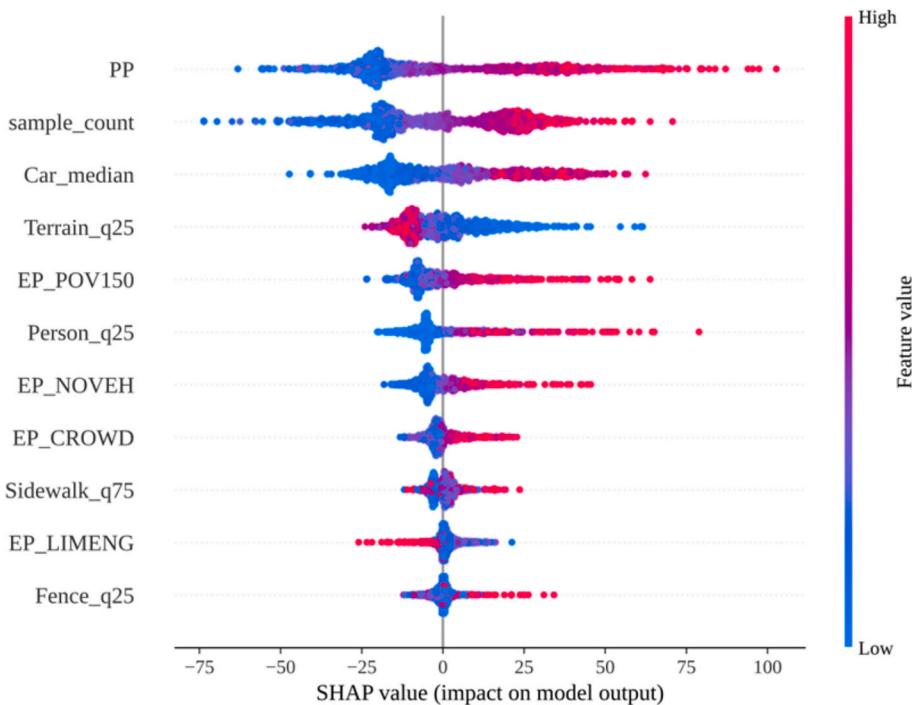


Fig. 14. SHAP summary plot of the modeling features on regional crash frequency.

elevated crash frequency. This trend likely reflects increased traffic volume or pedestrian activity along roads with these characteristics. Conversely, the 25 % percentile of terrain proportion in GSVs (Terrain_q25) demonstrated a negative impact on model predictions, with higher feature values (red points) predominantly associated with negative SHAP values, indicating that an increased terrain proportion can, in some instances, mitigate crash frequency. Regarding socio-demographic features, the proportion of persons below 150 % poverty (POV150), the proportion of households without vehicles (NOVEH), and the proportion of occupied housing units with more people than rooms (CROWD) all exhibit prominent positive effects on crash frequency. This pattern was particularly evident at higher feature values, where SHAP values were predominantly positive. These findings suggest that indicators of area vulnerability are positively correlated with local traffic safety issues. The strong positive effect of poverty rates on crash occurrences may be attributed to inadequate traffic infrastructure in economically disadvantaged areas. Similarly, a higher proportion of households without vehicles implies greater reliance on walking or cycling, potentially increasing exposure to traffic risks. The influence of overcrowded housing (CROWD) suggests that high housing density may contribute to increased traffic risks, possibly due to greater reliance on shared transportation or walking in these areas.

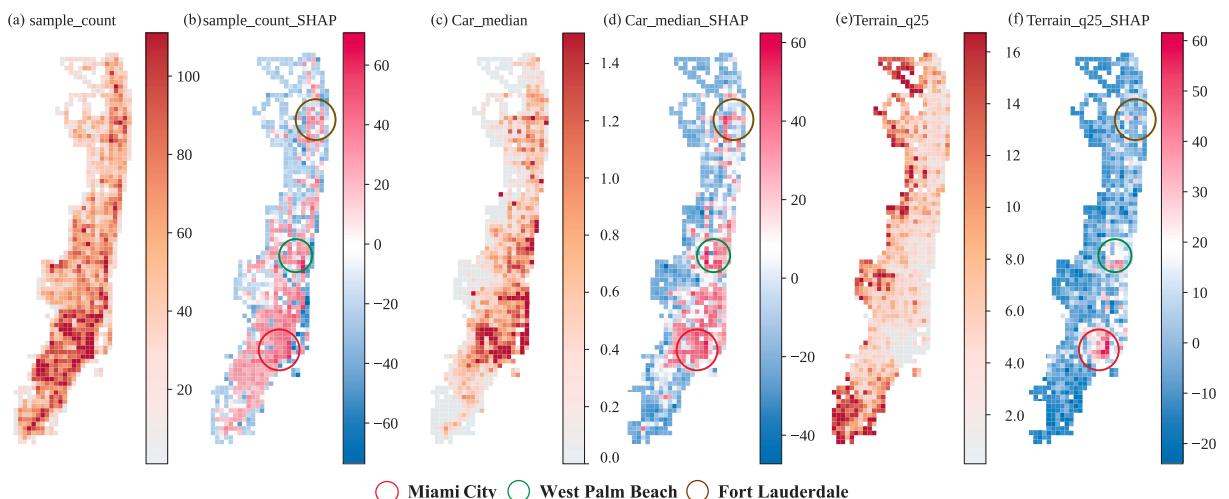


Fig. 15. Spatial distribution of key visual environment features (a, c, e) & their SHAP (b, d, f).

To further elucidate spatial patterns, Fig. 15 map the distribution of key visual environment features alongside their SHAP values. Sample count, representing the number of samples within the region road segments, exhibits a clear positive correlation with crash occurrence shown in Fig. 15 (b). This exposure variable positively impacts high-density areas, consistent with findings from previous research on exposure risk theory (Almasi and Behnood, 2022). The positive impact of the median car proportion in GSVs (Car_median) on regional crashes is predominantly observed in urban core areas shown in Fig. 15 (d), including the cities of Miami, West Palm Beach, and Fort Lauderdale. This indicates that the car proportion can serve as an effective indicator of local roadway traffic volume, significantly increasing the exposure of crashes in these densely populated urban areas. Notably, Fig. 15 (f) shows that an increase in terrain proportion along roads (Terrain_q25) is significantly associated with reduced crash frequency. On the one hand, areas with a higher terrain proportion are typically found in non-central or suburban regions as shown in Fig. 15 (e), which tend to have lower traffic density. These conditions contribute to reduce risky driving interactions and conflicts, thereby reducing the likelihood of traffic crashes in these areas. On the one hand, the presence of trees and shrubs on these terrain (greenery) can create the perception of narrower streets, prompting drivers to slow down and stay attentive to their surroundings (Abdel-Aty et al., 2024; Cai et al., 2022; Tice et al., 2024). Therefore, strategically adding visual elements such as trees and landscaping at current crash hotspots could be beneficial for enhancing road safety.

Furthermore, Fig. 16 presents the spatial distribution of key socio-demographic features, including total population (PP), proportion of persons below 150 % poverty (POV150), and proportion of households without vehicle (NOVEH), alongside the respective SHAP values. As depicted in Fig. 16 (b), total population (PP) shows a significant positive correlation with crash frequency, particularly in densely populated urban areas, indicating that higher population densities often result in elevated exposure to traffic risks (Buehler and Pucher, 2021; Papadimitriou et al., 2019). The proportion of persons below 150 % poverty (POV150), illustrated in Fig. 16 (c), reveals that regions with higher poverty levels, particularly in urban Miami and central Fort Lauderdale, are associated with greater crash risk shown in Fig. 16 (d). This pattern suggests that economically disadvantaged areas are more susceptible to traffic crashes due to inadequate infrastructure and lack of effective safety measures (Li et al., 2022; Liu and Sharma, 2018; Wulz et al., 2023). Likewise, proportion of households without vehicle (NOVEH) is also positively correlated with crash risk as shown in Fig. 16 (f). This trend is especially pronounced in the Miami urban core, where a higher proportion of vehicle-less households implies increased dependence on walking, cycling, and public transit, thereby raising exposure to traffic hazards. This trend is especially pronounced in the Miami urban core, where a higher proportion of vehicle-less households implies increased dependence on walking, cycling, and public transit, thereby raising exposure to traffic hazards. These findings are consistent with earlier studies indicating that residents in socio-economically disadvantaged neighborhoods experience higher crash frequency due to insufficient infrastructure and suboptimal traffic management (Kuo et al., 2024; Lee et al., 2014). Collectively, these results underscore the pressing need to address socio-economic disparities through targeted infrastructure development and enhanced safety measures to reduce crash risks in vulnerable communities.

Furthermore, we employed SHAP (SHapley Additive exPlanations) values to analyze the interaction effects of key features on regional traffic safety, focusing on three critical pairs: total population (PP) and median car proportion in GSVs (Car_median), terrain proportion along roads (Terrain_q25) and Car_median, as well as total population (PP) and proportion of persons below 150 % poverty (EP_POV150). The visualizations in Fig. 17 (left: scatter plots for SHAP dependency; right: contour plots for prediction interaction) reveal that each feature's contribution to crash prediction is conditional on interacting features. For instance, (1) the interaction analysis between total population (PP) and median car proportion in GSVs (Car_median) reveals that PP's SHAP values (indicating risk contribution) exhibit a significant increasing trend with the elevation of Car_median, uncovering a conditional effect: the amplifying impact of population density on regional traffic crash risk is more pronounced in areas with higher car density. Model prediction values

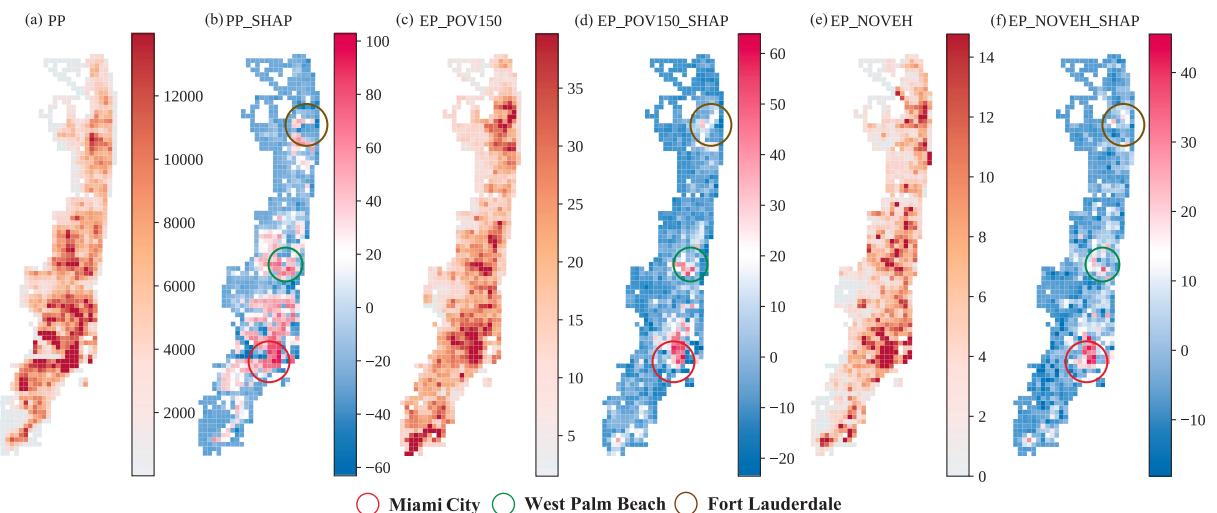


Fig. 16. Spatial distribution of key socio-demographic features (a, c, e) & their SHAP (b, d, f).

on the right further confirms that the positive interaction between these two variables significantly elevates regional traffic crash frequency; (2) The SHAP values of terrain proportion along roads (Terrain_q25) demonstrate a negatively correlated characteristic with median car proportion in GSVs (Car_median), meaning that as Car_median increases, the inhibitory effect of terrain proportion on regional crash risk strengthens. Model prediction results indicate that the feature combination of high Car_median and low Terrain_q25 corresponds to the highest crash frequency. This finding suggests that increasing green space or natural terrain proportion in high traffic density areas may mitigate traffic crash risk by improving road environmental characteristics; (3) Regarding the interaction effect between total population (PP) and proportion of persons below 150 % poverty (POV150), high joint values are significantly correlated with elevated SHAP values, highlighting the synergistic risk between socioeconomic factors and population characteristics: dense populations in low-income areas face more complex traffic crash threats, potentially related to the interplay of infrastructure levels, traffic management resources, and travel patterns. These results highlight the interdependent nature of features, forming complex risk landscapes unaccounted for by standalone analysis. Incorporating such interactions is crucial for model interpretability, revealing hidden relationships (e.g., poverty-population density hotspots) and enabling targeted interventions.

4.3. Case Study: Typical regional traffic safety scenarios

Previous results reveal significant variability in visual environment features and socio-demographic characteristics across different regions. To further illustrate their spatial heterogeneity on regional crash risk, detailed case studies were conducted. Fig. 18 presents a comparison of model prediction for two typical regional traffic safety scenarios, utilizing SHAP values to quantify the contribution of each feature to corresponding crashes. These two regions show significantly different visual environment and socio-demographic characteristics, resulting in pronounced differences in traffic crash frequency.

Fig. 18 (a) illustrates an urban area with the model predicts equaling 248.81, markedly higher than the baseline of 86.26. The results highlight that four visual environment features (Car_median, sample_count, Person_q25, and Terrain_q25) and three socio-demographic characteristics (PP, EP_NOVEH, EP_POV150) show positive effects (red SHAP arrows) in increasing the crash frequency. Notably, population size (PP), the proportion of cars in GSV (Car_median), and the number of GSV images (sample_count) are all above the average level. It indicates this area has a high population density, heavy vehicle volume, and dense road networks, therefore leading to a large number of crashes. Meanwhile, this region exhibits a higher proportion of person (Person_q25 = 0.011 %) and much less green space (Terrain_q25 = 0.9 %). These visual environment features, along with a high poverty rate (EP_POV150 = 34.43 %) and proportion of households without vehicles (EP_NOVEH = 15.69 %), appear to contribute significantly to the increased predicted values. In contrast, Fig. 18 (b) depicts a region with a low predicted crash frequency of 11.03. Besides its relatively low population size (PP = 871) and road density (sample count = 33), this area's visual environment is marked by a low proportion of cars (Car_median = 0.14 %), abundant green spaces (Terrain_q25 = 10.40 %), and no person along roads (Person_q25 = 0). All these factors have negative SHAP values as shown in Fig. 18 (b), suggesting that these visual environment features are associated to reduce the regional crash frequency. The socio-demographic characteristics also suggest improved living conditions and better income circumstances, such as lower poverty rates (EP_POV150 = 4.05 %) and a smaller proportion of households without vehicles (EP_NOVEH = 2.02 %). As a result, the model tends to predict lower overall values for these areas.

In general, we observe that various visual environment features significantly influence the predicted regional crash frequency. Dense urban environments with limited greenery and higher poverty levels are often linked to higher regional crash frequency, whereas areas featuring abundant green spaces and lower population densities tend to experience fewer traffic crashes. This case study underscores the model's sensitivity to visual environment features and suggests that these characteristics must be thoroughly considered when predicting and analyzing regional traffic safety issues.

4.4. Methodological reflections and future directions

From a modeling perspective, our work extends the Geographically Random Forest (GRF) (Georganos et al., 2021) framework by incorporating an adaptive multiscale bandwidth selection process, thereby calibrating each local model at its optimal spatial scale. This strategy is analogous to the transition from Geographically Weighted Regression (GWR) to Multiscale GWR (MGWR), in which MGWR permits different explanatory variables to operate at distinct spatial scales (Fotheringham et al., 2017). However, while MGWR's multiscale capability is confined to variable-specific bandwidths in a linear framework, our approach dynamically searches the spatial domain to determine the optimal scale for each local random forest.

In addition, our framework uniquely integrates SHAP values for model interpretability, thereby moving beyond conventional feature importance measures (Wang et al., 2024, 2024a; Wu et al., 2024). By quantifying the contribution of each predictor to the model output, SHAP enhances our understanding of the complex, nonlinear interactions underlying crash data, thereby improving interpretability without sacrificing predictive performance.

In our current implementation, we employ out-of-bag mean squared error (OOB MSE) as the default optimization metric. Nevertheless, alternative metrics such as Moran's I can offer deeper insights into spatial correlation (Georganos et al., 2021). Our preliminary experiments indicate that while optimizing based on Moran's I provides a more rigorous measure of spatial dependence, it significantly increases training time—approximately threefold compared to OOB MSE—thus presenting a trade-off between computational efficiency and the extent of spatial characterization. Future research should explore hybrid or fully adaptive optimization strategies that combine multiple criteria, such as block cross-validation and AICs (Merola, 2023; Wagenmakers and Farrell, 2004), to better balance predictive accuracy and spatial consistency.

Moreover, although our implementation currently relies on random forests, the MGRF framework is inherently flexible and can be

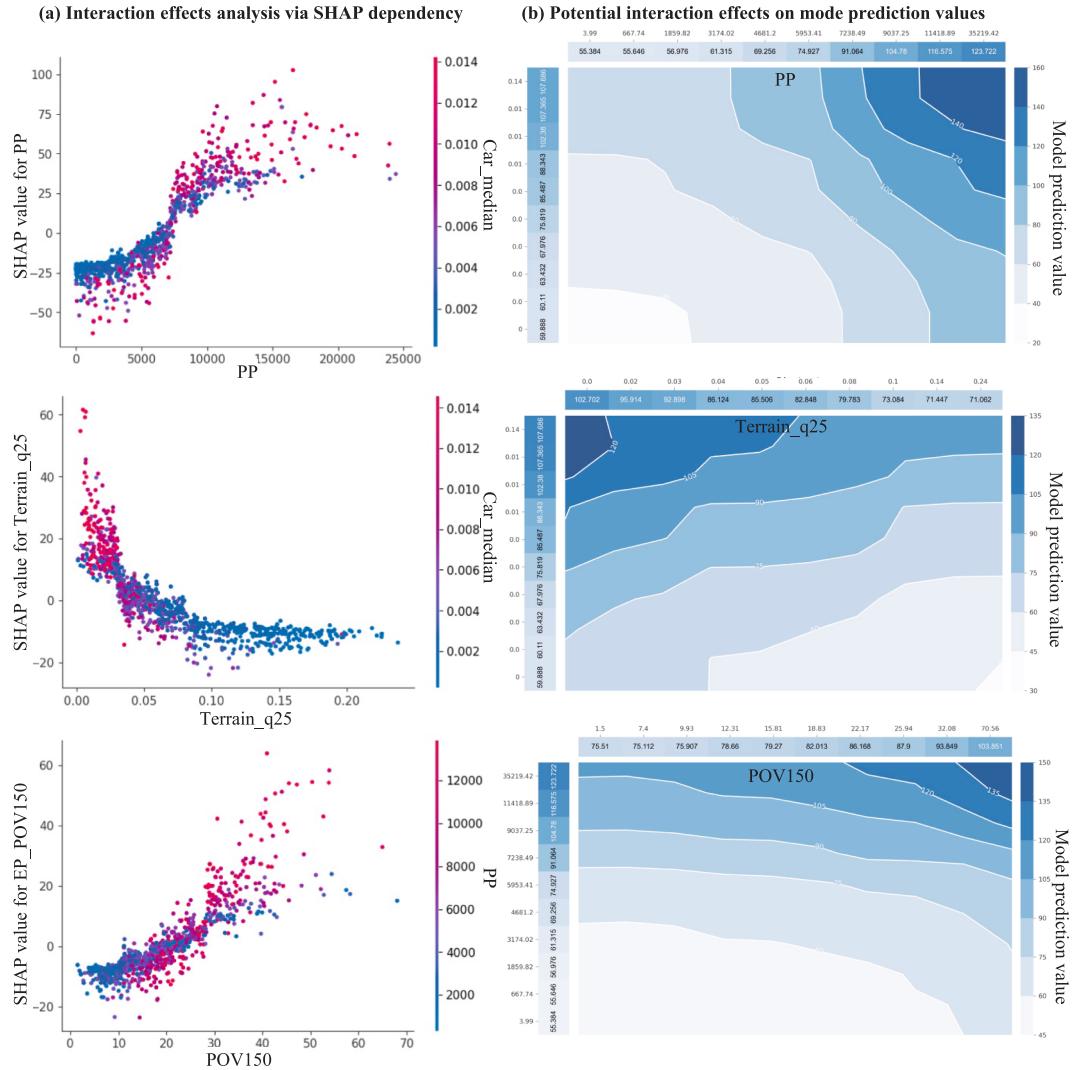


Fig. 17. Interaction effects of key features on regional traffic safety via SHAP dependency and model prediction values.

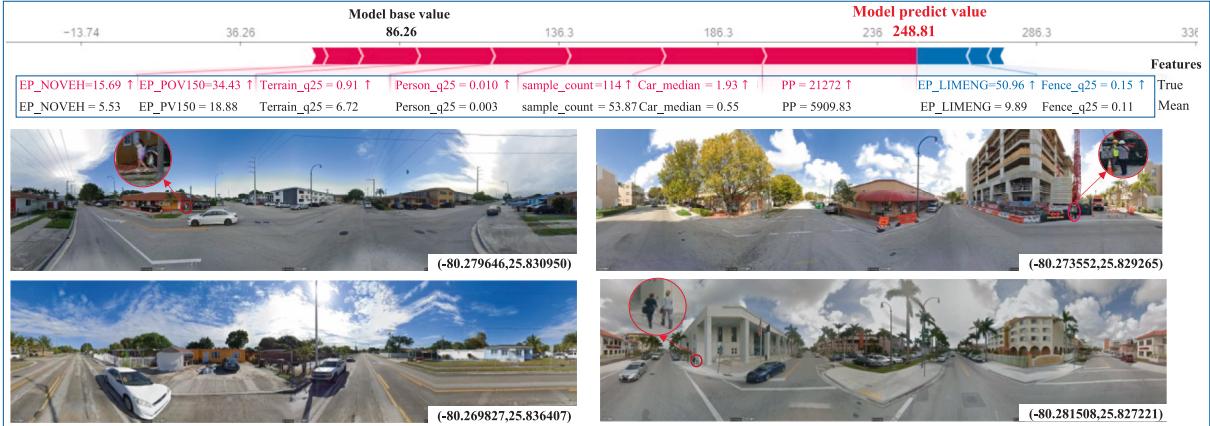
extended to incorporate alternative machine learning algorithms, such as eXtreme Gradient Boosting (XGBoost), Bayesian additive regression trees (BART), or multilayer perceptron (MLP). While these alternatives may further enhance predictive performance and broaden the model's applicability, they may also introduce additional challenges in hyperparameter tuning and model interpretability. These methodological refinements represent promising avenues for future investigation and hold the potential to contribute significantly to the broader fields of spatial analysis and machine learning.

5. Conclusions

Traffic crashes remain a critical concern for public safety and urban planning. To improve safety, macro-level traffic safety modeling was proposed to identify risky factors to crashes. Recently, spatial statistical and ML models have been developed to uncover safety-critical features including socio-demographics (Lee et al., 2015; Tang et al., 2020), road network configuration (Huang et al., 2010; Merlin et al., 2020), and travel patterns (Lee et al., 2014). However, two major challenges still need to be addressed: (1) Existing features mainly focus on macro-level, static infrastructure and traffic conditions, ignoring drivers' visual perception of their surrounding environment (Abdel-Aty et al., 2024; Merlin et al., 2020). (2) Current spatial models assume spatial impacts to be either homogeneous or varying at a fixed spatial scale (Fotheringham et al., 2017), limiting their ability to handle the inherent multiscale and dynamic spatial heterogeneity between realistic traffic features and crashes.

To address the abovementioned issues, we utilized advanced image semantic segmentation techniques to extract the visual environment features (e.g., cars, buildings) and integrated them with socio-demographic features (e.g., population) for macro-level crash prediction. To overcome the fixed spatial scale limitation, we introduced a novel spatial modeling approach MGRF to

(a). High crash risk region (Total crash frequency = 241)



(b). Low crash risk region (Total crash frequency = 10)



Fig. 18. Evidence of visual environment and socio-demographic characteristics on crashes: visualization of SHAP values with location context.

adaptively fit the optimal spatial bandwidth, which enable it to conduct multiscale spatial modeling on capturing spatial heterogeneity. Data from Southeast Florida (Miami-Dade, Broward, and Palm Beach counties) were utilized for empirical experiments. The results shows that the proposed MGRF achieved better model predictions and identified most important risk features and their spatial heterogeneity. Overall, the main findings of the study can be summarized as:

- 1) MGRF can adaptively capture the spatial variation of feature impacts and responds more effectively to spatial heterogeneity across different regions, significantly improving the accuracy of regional crash predictions. Compared to exiting models, MGRF achieves 30.31 %, 9.98 %, and 5.83 % improvements in MSE, MAE, and R², respectively. It further outperforms the fixed-scale GRF model by 9.67 % in MSE, 3.03 % in MAE, and 1.54 % in R².
- 2) The introduction of visual environment features led to a substantial 15.08 % improvement in R², from of 0.7354 to 0.8462, highlighting the critical role of these features in macro-level crash modeling. Empirical results reveal that dense urban areas with low green space and a higher proportion of vehicles were associated with increased crash frequency. In contrast, regions with abundant green space and sparse road density exhibited reduced crashes, suggesting that roadside green environments may enhance driving attention to be benefit for improving safety.
- 3) The relationship between socio-demographic factors—including poverty rate, population density, along with vehicle ownership—reveals pronounced spatial heterogeneity in traffic crash frequency across regions. Findings indicate that in densely populated areas, high poverty rates and low vehicle ownership are frequently associated with higher traffic crash frequencies, underscoring the need for targeted traffic safety interventions that account for region-specific risk factors.

To sum up, this study supports the expanded use of visual environment features in macro-level crash prediction models, offering a promising direction for enhancing the accuracy and interpretability of traffic safety evaluations. By capturing both the local and global spatial variations in crash determinants, the MGRF model provides a more robust foundation for future crash risk analysis. Nonetheless, there are still a few limitations in the current study. First, the current study was based on a specific region, which may limit the model's transferability to other areas. Future research should consider testing the model in diverse geographic contexts to evaluate its generalizability. Second, while the model incorporates visual and socio-demographic features, additional factors such as weather

conditions, traffic volume, and road quality could further enhance prediction accuracy. Finally, future work should explore hybrid or fully adaptive optimization strategies that more seamlessly integrate spatial dependence while maintaining computational efficiency.

CRediT authorship contribution statement

Pengfei Cui: Writing – review & editing, Formal analysis, Methodology, Data curation, Writing – original draft, Conceptualization. **Mohamed Abdel-Aty:** Validation, Project administration, Formal analysis, Writing – review & editing, Investigation, Conceptualization, Writing – original draft, Supervision, Funding acquisition. **Lei Han:** Writing – review & editing, Investigation, Conceptualization, Writing – original draft, Data curation, Methodology, Formal analysis. **Xiaobao Yang:** Writing – review & editing, Funding acquisition, Supervision, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Key R&D Program of China (No. 2023YFC3009600), the Key Program of the National Natural Science Foundation of China (No. 62333016). The authors wish to thank Dr. Dongdong Wang for his great support for data preparation in this study and the contribution of the UCF's Smart and Safe Transportation lab.

REFERENCES

- Abdel-Aty, M., Ugan, J., Islam, Z., 2024. Exploring the influence of drivers' visual surroundings on speeding behavior. *Accid. Anal. Prev.* 198, 107479. <https://doi.org/10.1016/j.aap.2024.107479>.
- Almasi, S.A., Behnood, H.R., 2022. Exposure based geographic analysis mode for estimating the expected pedestrian crash frequency in urban traffic zones; case study of Tehran. *Accid. Anal. Prev.* 168, 106576. <https://doi.org/10.1016/j.aap.2022.106576>.
- Azimian, A., Dimitra Pyrialakou, V., Lavrenz, S., Wen, S., 2021. Exploring the effects of area-level factors on traffic crash frequency by severity using multivariate space-time models. *Anal. Methods Accid. Res* 31, 100163. <https://doi.org/10.1016/j.amar.2021.100163>.
- Barua, S., El-Basyouny, K., Islam, M.T., 2016. Multivariate random parameters collision count data models with spatial heterogeneity. *Anal. Methods Accid. Res* 9, 1–15. <https://doi.org/10.1016/j.amar.2015.11.002>.
- Bivand, R., Müller, W.G., Reder, M., 2009. Power calculations for global and local Moran's I. *Comput. Stat. Data Anal.* 53, 2859–2872. <https://doi.org/10.1016/j.csda.2008.07.021>.
- Breiman, L., 2001. Random Forests. *Mach. Learn.* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Brunsdon, C., Fotheringham, S., Charlton, M., 1998. Geographically Weighted Regression. *J. r. Stat. Soc. Ser. Stat.* 47, 431–443. <https://doi.org/10.1111/1467-9884.00145>.
- Buehler, R., Pucher, J., 2021. The growing gap in pedestrian and cyclist fatality rates between the United States and the United Kingdom, Germany, Denmark, and the Netherlands, 1990–2018. *Transp. Rev.* 41, 48–72. <https://doi.org/10.1080/01441647.2020.1823521>.
- Cai et al. - 2022 - Applying machine learning and google street view t.pdf, n.d.
- Cai, Q., Abdel-Aty, M., Yuan, J., Lee, J., Wu, Y., 2020. Real-time crash prediction on expressways using deep generative models. *Transp. Res. Part C Emerg. Technol.* 117, 102697. <https://doi.org/10.1016/j.trc.2020.102697>.
- Cai, Q., Abdel-Aty, M., Zheng, O., Wu, Y., 2022. Applying machine learning and google street view to explore effects of drivers' visual environment on traffic safety. *Transp. Res. Part C Emerg. Technol.* 135, 103541. <https://doi.org/10.1016/j.trc.2021.103541>.
- Cui, P., Abdel-Aty, M., Wang, C., Yang, X., Song, D., 2025. Examining the impact of spatial inequality in socio-demographic and commute patterns on traffic crash rates: Insights from interpretable machine learning and spatial statistical models. *Transp. Policy* 167, 222–245. <https://doi.org/10.1016/j.tranpol.2025.03.033>.
- Cui, P., Abdel-Aty, M., Yang, X., Wang, C., Yuan, Y., 2025. Quantifying spatial inequities in traffic injury rates through the integration of urban road network measures and social vulnerability. *Accid. Anal. Prev.* 211, 107916. <https://doi.org/10.1016/j.aap.2025.107916>.
- Cui, P., Yang, X., Abdel-Aty, M., Zhang, J., Yan, X., 2024. Advancing urban traffic accident forecasting through sparse spatio-temporal dynamic learning. *Accid. Anal. Prev.* 200, 107564. <https://doi.org/10.1016/j.aap.2024.107564>.
- Deng, L., Adjoudi, M., Rishe, N., 2020. Inverse Distance Weighted Random Forests: Modeling Unevenly distributed Non-Stationary Geographic Data. In: In: 2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS). Presented at the 2020 International Conference on Advanced Computer Science and Information Systems (ICACSIS), pp. 41–46. <https://doi.org/10.1109/ICACSIS51025.2020.9263208>.
- Department for Transport, 2023. Reported road casualties Great Britain, annual report: 2022. UK, GOV.
- Drägut, L., Eisank, C., Strasser, T., 2011. Local variance for multi-scale analysis in geomorphometry. *Geomorphology* 130, 162–172. <https://doi.org/10.1016/j.geomorph.2011.03.011>.
- Ewing, R., Hamidi, S., Grace, J.B., 2016. Urban sprawl as a risk factor in motor vehicle crashes. *Urban Stud.* 53, 247–266. <https://doi.org/10.1177/0042098014562331>.
- Fan, C., Xu, J., Natarajan, B.Y., Mostafavi, A., 2023a. Interpretable machine learning learns complex interactions of urban features to understand socio-economic inequality. *Comput. Civ. Infrastruct. Eng.* 38, 2013–2029.
- Fan, Z., Zhang, F., Loo, B.P.Y., Ratti, C., 2023b. Urban visual intelligence: Uncovering hidden city profiles with street view images. *Proc. Natl. Acad. Sci.* 120, e2220417120. <https://doi.org/10.1073/pnas.2220417120>.
- Fotheringham, A.S., Charlton, M., Brunsdon, C., 1997. Measuring Spatial Variations in Relationships with Geographically Weighted Regression. In: Fischer, M.M., Getis, A. (Eds.), Recent Developments in Spatial Analysis: Spatial Statistics, Behavioural Modelling, and Computational Intelligence. Springer, Berlin, Heidelberg, pp. 60–82. https://doi.org/10.1007/978-3-662-03499-6_4.
- Fotheringham, A.S., Yang, W., Kang, W., 2017. Multiscale Geographically Weighted Regression (MGWR). *Ann. Am. Assoc. Geogr.* 107, 1247–1265. <https://doi.org/10.1080/24694452.2017.1352480>.
- Georganos, S., Grippa, T., Niang Gadiaga, A., Linard, C., Lenert, M., Vanhuyse, S., Mboga, N., Wolff, E., Kalogirou, S., 2021. Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. *Geocarto Int.* 36, 121–136. <https://doi.org/10.1080/10106049.2019.1595177>.
- Gu, Y., Liu, D., Arvin, R., Khattak, A.J., Han, L.D., 2023. Predicting intersection crash frequency using connected vehicle data: a framework for geographical random forest. *Accid. Anal. Prev.* 179, 106880. <https://doi.org/10.1016/j.aap.2022.106880>.

- Han, L., Abdel-Aty, M., 2025. Intersection crash analysis considering longitudinal and lateral risky driving behavior from connected vehicle data: A spatial machine learning approach. *Accid. Anal. Prev.* 220, 108180.
- Han, L., Abdel-Aty, M., Joo, Y. J., Zhai, S., & Wang, D., 2025. Intersection Crash Frequency Analysis Considering Visual Environment Features Using Random Parameter Negative Binomial-Lindley Model. *Transportation Research Record*, 03611981251353186.
- Han, L., Abdel-Aty, M., Yu, R., Wang, C., 2024a. LSTM + Transformer Real-Time Crash Risk Evaluation using Traffic Flow and Risky Driving Behavior Data. *IEEE Trans. Intell. Transp. Syst.* 1–13. <https://doi.org/10.1109/TITS.2024.3438616>.
- Han, L., Yu, R., Wang, C., Abdel-Aty, M., 2024b. Transformer-based modeling of abnormal driving events for freeway crash risk evaluation. *Transp. Res. Part C Emerg. Technol.* 165, 104727. <https://doi.org/10.1016/j.trc.2024.104727>.
- Harris, N.L., Goldman, E., Gabris, C., Nordling, J., Minnemeyer, S., Ansari, S., Lippmann, M., Bennett, L., Raad, M., Hansen, M., 2017. Using spatial statistics to identify emerging hot spots of forest loss. *Environ. Res. Lett.* 12, 024012. <https://doi.org/10.1088/1748-9326/aa5a2f>.
- Hong, J., Shankar, V.N., Venkataraman, N., 2016. A spatially autoregressive and heteroskedastic space-time pedestrian exposure modeling framework with spatial lags and endogenous network topologies. *Anal. Methods Accid. Res* 10, 26–46. <https://doi.org/10.1016/j.amar.2016.05.001>.
- Hou, W., Li, D., Xu, C., Zhang, H., Li, T., 2018. An advanced k nearest neighbor classification algorithm based on KD-tree. In: In: 2018 IEEE International Conference of Safety Produce Informatization (IICSP). IEEE, pp. 902–905. <https://doi.org/10.1109/IICSP18.8690508>.
- Huang, H., Abdel-Aty, M.A., Darwiche, A.L., 2010. County-level crash risk analysis in Florida: Bayesian spatial modeling. *Transp. Res.* 2148, 27–37. <https://doi.org/10.3141/2148-04>.
- Kim, Y., Kim, Y., 2022. Explainable heat-related mortality with random forest and SHapley Additive exPlanations (SHAP) models. *Sustain. Cities Soc.* 79, 103677. <https://doi.org/10.1016/j.scs.2022.103677>.
- Krueger, R., Bansal, P., Buddhabarapu, P., 2020. A new spatial count data model with Bayesian additive regression trees for accident hot spot identification. *Accid. Anal. Prev.* 144, 105623. <https://doi.org/10.1016/j.aap.2020.105623>.
- Kuo, P.-F., Lord, D., Sulistyah, U.D., Jhamb, A., 2024. The impact of the great recession on the spatial patterns of traffic fatalities in Texas: a spatial point pattern test. *Transp. Transp. Sci.* 20, 2145861. <https://doi.org/10.1080/23249935.2022.2145861>.
- Lee, J., Abdel-Aty, M., Choi, K., 2014. Analysis of residence characteristics of at-fault drivers in traffic crashes. *Saf. Sci.* 68, 6–13. <https://doi.org/10.1016/j.ssci.2014.02.019>.
- Lee, J., Abdel-Aty, M., Jiang, X., 2015. Multivariate crash modeling for motor vehicle and non-motorized modes at the macroscopic level. *Accid. Anal. Prev.* 78, 146–154. <https://doi.org/10.1016/j.aap.2015.03.003>.
- Lee, J., Liu, H., Abdel-Aty, M., 2023. Changes in traffic crash patterns: before and after the outbreak of COVID-19 in Florida. *Accid. Anal. Prev.* 190, 107187. <https://doi.org/10.1016/j.aap.2023.107187>.
- Li, X., Yu, S., Huang, X., Dadashova, B., Cui, W., Zhang, Z., 2022. Do underserved and socially vulnerable communities observe more crashes? a spatial examination of social vulnerability and crash risks in Texas. *Accid. Anal. Prev.* 173, 106721. <https://doi.org/10.1016/j.aap.2022.106721>.
- Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., Zhang, W., 2015. Assessing street-level urban greenery using Google Street View and a modified green view index. *Urban For. Urban Green.* 14, 675–685. <https://doi.org/10.1016/j.ufug.2015.06.006>.
- Lin, L., Shi, F., Li, W., 2021. Assessing inequality, irregularity, and severity regarding road traffic safety during COVID-19. *Sci. Rep.* 11, 13147. <https://doi.org/10.1038/s41598-021-91392-z>.
- Liu, C., Sharma, A., 2018. Using the multivariate spatio-temporal Bayesian model to analyze traffic crashes by severity. *Anal. Methods Accid. Res* 17, 14–31. <https://doi.org/10.1016/j.amar.2018.02.001>.
- Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2020. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* 2, 56–67. <https://doi.org/10.1038/s42256-019-0138-9>.
- Lundberg, S.M., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions, in: Advances in Neural Information Processing Systems. Curran Associates, Inc.
- Mannering, F.L., Bhat, C.R., 2014. Analytic methods in accident research: Methodological frontier and future directions. *Anal. Methods Accid. Res* 1, 1–22. <https://doi.org/10.1016/j.amar.2013.09.001>.
- Merlin, L.A., Guerra, E., Dumbaugh, E., 2020. Crash risk, crash exposure, and the built environment: a conceptual review. *Accid. Anal. Prev.* 134, 105244. <https://doi.org/10.1016/j.aap.2019.07.020>.
- Merola, G.M., 2023. Blocked Cross-Validation: A Precise and Efficient Method for Hyperparameter Tuning. <https://doi.org/10.48550/arXiv.2306.06591>.
- Mukaka, M., 2012. A guide to appropriate use of Correlation coefficient in medical research. *Malawi Med. J. J. Med. Assoc. Malawi* 24, 69–71.
- Naznin, F., Currie, G., Logan, D., Sarvi, M., 2016. Application of a random effects negative binomial model to examine tram-involved crash frequency on route sections in Melbourne, Australia. *Accid. Anal. Prev.* 92, 15–21. <https://doi.org/10.1016/j.aap.2016.03.012>.
- NHTSA, 2023. Traffic safety facts: Crash Stats. Natl. Highw. Traffic Saf. Adm. Washington DC DOT HS 813 530.
- Odijk, M.J.M., Asadi, M., Baran Ulak, M., Geurs, K.T., 2023. The interactions between accessibility and crash risk from a social equity perspective: a case study at the Rotterdam-the Hague metropolitan region. *J. Saf. Res.* 87, 176–186. <https://doi.org/10.1016/j.jsr.2023.09.015>.
- O'Sullivan, D., 2003. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships* (review). *Geogr. Anal.* 35, 272–275.
- Papadimitriou, E., Filtness, A., Theofilatos, A., Ziakopoulos, A., Quigley, C., Yannis, G., 2019. Review and ranking of crash risk factors related to the road infrastructure. *Accid. Anal. Prev.* 125, 85–97. <https://doi.org/10.1016/j.aap.2019.01.002>.
- Rahman, M.H., Antipova, A., 2024. Structural equation model in exploring urban sprawl and its impact on commuting time in 162 US urbanized areas. *Cities* 148, 104855. <https://doi.org/10.1016/j.cities.2024.104855>.
- Ren, H., Song, Y., Wang, J., Hu, Y., Lei, J., 2018. In: A Deep Learning Approach to the Citywide Traffic Accident Risk Prediction, in, pp. 3346–3351. <https://doi.org/10.1109/ITSCC.2018.8569437>.
- Ribeiro, M.T., Singh, S., Guestrin, C., 2016. In: "why Should I Trust You?": Explaining the Predictions of Any Classifier, in. Association for Computing Machinery, New York, NY, USA, pp. 1135–1144. <https://doi.org/10.1145/2939672.2939778>.
- Siddiqui, C., Abdel-Aty, M., Choi, K., 2012. Macroscopic spatial analysis of pedestrian and bicycle crashes. *Accid. Anal. Prev.* 45, 382–391. <https://doi.org/10.1016/j.aap.2011.08.003>.
- Speiser, J.L., Miller, M.E., Tooze, J., Ip, E., 2019. A comparison of random forest variable selection methods for classification prediction modeling. *Expert Syst. Appl.* 134, 93–101. <https://doi.org/10.1016/j.eswa.2019.05.028>.
- Strudel, R., Garcia, R., Laptev, I., Schmid, C., 2021. Segmenter: Transformer for Semantic Segmentation. In: Presented at the Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 7262–7272.
- Strumbelj, E., Kononenko, I., 2014. Explaining prediction models and individual predictions with feature contributions. *Knowl. Inf. Syst.* 41, 647–665. <https://doi.org/10.1007/s10115-013-0679-x>.
- Sun, K., Zhou, R.Z., Kim, J., Hu, Y., n.d. PyGRF: An Improved Python Geographical Random Forest Model and Case Studies in Public Health and Natural Disasters. *Trans. GIS* n/a. Doi: 10.1111/tgis.13248.
- Sun, Z., Wang, D., Gu, X., Abdel-Aty, M., Xing, Y., Wang, J., Lu, H., Chen, Y., 2023. A hybrid approach of random forest and random parameters logit model of injury severity modeling of vulnerable road users involved crashes. *Accid. Anal. Prev.* 192, 107235. <https://doi.org/10.1016/j.aap.2023.107235>.
- Tang, J., Gao, F., Liu, F., Han, C., Lee, J., 2020. Spatial heterogeneity analysis of macro-level crashes using geographically weighted Poisson quantile regression. *Accid. Anal. Prev.* 148, 105833. <https://doi.org/10.1016/j.aap.2020.105833>.
- Tang, X., Bi, R., Wang, Z., 2023. Spatial analysis of moving-vehicle crashes and fixed-object crashes based on multi-scale geographically weighted regression. *Accid. Anal. Prev.* 189, 107123. <https://doi.org/10.1016/j.aap.2023.107123>.
- Tice, P.C., dey Tirtha, S., Eluru, N., Hancock, P.A., 2024. Attentiveness in urban spaces: The rhythm of the street. *Transp. Res. Interdiscip. Perspect.* 25, 101063. Doi: 10.1016/j.trip.2024.101063.
- Trirat, P., Yoon, S., Lee, J.-G., 2023. MG-TAR: Multi-View Graph Convolutional Networks for Traffic Accident Risk Prediction. *IEEE Trans. Intell. Transp. Syst.* 24, 3779–3794. <https://doi.org/10.1109/TITS.2023.3237072>.

- Wagenmakers, E.-J., Farrell, S., 2004. AIC model selection using Akaike weights. *Psychon. Bull. Rev.* 11, 192–196. <https://doi.org/10.3758/BF03206482>.
- Wang, C., Abdel-Aty, M., Han, L., 2025. Grouped random parameters Poisson-Lindley model with spatial effects addressing crashes at intersections: Insights from visual environment features and spatiotemporal instability. *Analytic Methods in Accident Research*, 100387.
- Wang, C., Abdel-Aty, M., Cui, P., Han, L., 2024a. Effects of helmet usage on moped riders' injury severity in moped-vehicle crashes: Insights from partially temporal constrained random parameters bivariate probit models. *Accid. Anal. Prevent.* 208, 107800. <https://doi.org/10.1016/j.aap.2024.107800>.
- Wang, C., Abdel-Aty, M., Han, L., 2024b. Effects of speed difference on injury severity of freeway rear-end crashes: Insights from correlated joint random parameters bivariate probit models and temporal instability. *Anal. Methods Acc. Res.* 42, 100320.
- Wang, S., Gao, K., Zhang, L., Yu, B., Easa, S.M., 2024. Geographically weighted machine learning for modeling spatial heterogeneity in traffic crash frequency and determinants in US. *Accid. Anal. Prev.* 199, 107528. <https://doi.org/10.1016/j.aap.2024.107528>.
- Wen, X., Xie, Y., Jiang, L., Li, Y., Ge, T., 2022. On the interpretability of machine learning methods in crash frequency modeling and crash modification factor development. *Accid. Anal. Prev.* 168, 106617. <https://doi.org/10.1016/j.aap.2022.106617>.
- Wen, X., Xie, Y., Jiang, L., Pu, Z., Ge, T., 2021. Applications of machine learning methods in traffic crash severity modelling: current status and future directions. *Transp. Rev.* 41, 855–879. <https://doi.org/10.1080/01441647.2021.1954108>.
- WHO, 2022. Regional approach to the decade of action for road safety 2021–2030.
- Wu, D., Zhang, Y., Xiang, Q., 2024. Geographically weighted random forests for macro-level crash frequency prediction. *Accid. Anal. Prev.* 194, 107370. <https://doi.org/10.1016/j.aap.2023.107370>.
- Wu, J., Jia, P., Feng, T., Li, H., Kuang, H., 2023a. Spatiotemporal analysis of built environment restrained traffic carbon emissions and policy implications. *Transp. Res. Part Transp. Environ.* 121, 103839. <https://doi.org/10.1016/j.trd.2023.103839>.
- Wu, P., Chen, T., Diew Wong, Y., Meng, X., Wang, X., Liu, W., 2023b. Exploring key spatio-temporal features of crash risk hot spots on urban road network: a machine learning approach. *Transp. Res. Part Policy Pract.* 173, 103717. <https://doi.org/10.1016/j.tra.2023.103717>.
- Wulz, A.R., Sharpe, J.D., Miller, G.F., Wolkin, A.F., 2023. Association between social vulnerability factors and unintentional fatal injury rates—United States, 2015–2019. *J. Saf. Res.* 86, 245–252. <https://doi.org/10.1016/j.jsr.2023.07.003>.
- Yang, J., Zhao, L., McBride, J., Gong, P., 2009. Can you see green? Assessing the visibility of urban forests in cities. *Landsc. Urban Plan.* 91, 97–104. <https://doi.org/10.1016/j.landurbplan.2008.12.004>.
- Yocum, R.L., Gayah, V.V., 2022. County-level crash prediction models for Pennsylvania accounting for income characteristics. *Transp. Res. Interdiscip. Perspect.* 13, 100562. <https://doi.org/10.1016/j.trip.2022.100562>.
- Yu, B., Chen, Y., Bao, S., 2019. Quantifying visual road environment to establish a speeding prediction model: an examination using naturalistic driving data. *Accid. Anal. Prev.* 129, 289–298. <https://doi.org/10.1016/j.aap.2019.05.011>.
- Yu, R., Abdel-Aty, M., 2013. Utilizing support vector machine in real-time crash risk evaluation. *Accid. Anal. Prev.* 51, 252–259. <https://doi.org/10.1016/j.aap.2012.11.027>.
- Yu, R., Han, L., Zhang, H., 2021. Trajectory data based freeway high-risk events prediction and its influencing factors analyses. *Accid. Anal. Prev.* 154, 106085. <https://doi.org/10.1016/j.aap.2021.106085>.
- Yu, R., He, Y., Li, H., Li, S., Jian, B., 2024. RiskFormer: Exploring the temporal associations between multi-type aberrant driving events and crash occurrence. *Accid. Anal. Prev.* 206, 107698. <https://doi.org/10.1016/j.aap.2024.107698>.
- Zhao, S., Khattak, A.J., 2018. Injury severity in crashes reported in proximity of rail crossings: the role of driver inattention. *J. Transp. Saf. Secur.* 10, 507–524. <https://doi.org/10.1080/19439962.2017.1323061>.
- Zheng, Q., Sharmin, F., Xu, C., Liu, P., 2024. Assessing regional road traffic safety in Sweden through dynamic panel data analysis: Influence of the planned innovative policies and the unplanned COVID-19 pandemic. *Transp. Res. Part Policy Pract.* 179, 103918. <https://doi.org/10.1016/j.tra.2023.103918>.
- Zhou, D., Gayah, V.V., Wood, J.S., 2023. Integration of machine learning and statistical models for crash frequency modeling. *Transp. Lett.* 15, 1408–1419. <https://doi.org/10.1080/19427867.2022.2158257>.