**Identifying Traffic Accident Black Spots Using Built Environment Features and Machine Learning: A Case Study of Camden**

**Abstract**

Road traffic accidents have long-term been the main safety threat to urban residents. Identifying the high-incidence areas of traffic accidents, the traffic black spots, are of great significance in improving the traffic safety level in the road environment.

The study outlines an attempt to train random forest classification models to identify whether a junction is a potential traffic black spot, in the Camden borough of London. 4 dimensions of built environment features, including street view information, urban activities, junction structure and facilities, and road network configurations, are considered as independent features.

Around the accident risk of junctions, the trained model present a high overall accuracy of 0.71. In terms of the feature importance, whether the junction is formed based on only minor roads, and the multi-scale road network configuration features, play the most important roles in the classification model. The street view information of the road junctions, though tested effective in identifying traffic black spots in previous studies, have relatively poor performance in predicting and classifying junctions with different risk levels.

According to the study, the risk levels of road junctions in Camden can be regarded as more a global feature rather than a local feature related to the urban built environment. This research provides a reference for the management and maintenance of current road junctions, and for the design of safer road network system in the future.

**Keywords**

Road Accidents, Road Junction, Random Forest Classifier, Street View, Space Syntax

# Introduction

* History of the broader field

Road traffic accidents have long-term been the leading safety threat to urban residents. According to the statistics of the World Health Organization (WHO), there is an estimate of 1.35 million people dead in road traffic accidents every year. In high-income countries, the average death rate is 8.3 per 100,000 population(World Health Organization, 2018).

* Main Concept

Due to the huge hazard of traffic accidents, notice boards with black dots were set on road sections to warn drivers of traffic risks, as an accident prevention measurement(Leslie, 1977). The traffic black spots hence refer to concentration areas or high-risk locations of traffic accidents on the road (Dereli & Erdogan, 2017; Cui et al., 2022). Considering that, on a global scale, cities adopt similar forms of traffic organisation and more or less face the problem of reducing traffic risks, understanding the common causes of traffic black spots and exploring universal methods for identifying and predicting traffic black spots will have outstanding significance.

* Gap

However, currently, more researches focus on the risk (Ren et al., 2017) and the severity (Wang & Kockelman, 2013; Chang et al., 2022) of traffic accidents, while the traffic black spots are of limited concern. Some studies have explored the built environment triggers of traffic accidents (Ewing & Cervero, 2010; Asadi et al., 2022), but few have discussed the structural causes of traffic black spots. In addition, some studies have tried to accurately identify traffic black spots based on historical traffic accident data. Some advanced spatial analysis models and statistical methods have been proposed (Zhang, Shu & Yan, 2019; Yuan, Zeng & Shi, 2020). However, due to the poor openness of location-labelled traffic accident data between countries and regions, the cognition and identification of traffic black spots are often limited to the local areas.

* Opportunities

The global popularity of open-source urban built environment data, such as Point of Interest(POI) and Street View Images(SVI), and the growing accessibility of various machine learning models further enhances people's ability to mine and process urban information, and provides a rich and stable information source for understanding urban activities. This inspires a new possibility to explore a universal method for understanding and predicting traffic black spots.

* Task

The study outlines a traffic black spot detection method based on open-source urban built environment data. Specifically, considering that traffic accidents in the urban environment is prone to happen on the road junctions (Ewing and Dumbaugh, 2009；Siddiqui, Abdel-Aty and Choi, 2012；Kullgren et al., 2019)，the study develops a predictive framework to identify whether a road junction is a potential traffic black spot, based on 4 dimensions of built environment features on the junctions. Selected feature dimensions include, the structure and facility info of junctions, the road network configuration, the visual elements of the junction from street view, and the commercial and public activities around the junction. All the features are collected from open-source datasets with high transparency.

* Questions

Two key questions the research hope to answer are:

(1) To what extent, the static built environment feautres can be used to predict the risk of road junctions as traffic black spots?

(2) Which built environment features, compared to other features may contribute more to the accurate prediction of risk as traffice black spot for road junctions.

# Literature Review

## Traffic Accidents and Built Environment Triggers

Numerous studies have explored and examined the built-environment triggers for traffic accidents. In terms of accident types, most studies focus on road crashes dominated by vehicle traffic, and the common types of accidents include vehicle-to-vehicle (Asadi et al., 2022), vehicle-to-pedestrian (Ukkusuri et al., 2012; Osama & Sayed, 2017), and vehicle-to-bicycle (Gitelman et al., 2017).

As for the built-environment triggers, in a broad sense, the built environment features can be associated with all physical structures built by humans to support human activities, covering cities, villages, buildings, infrastructures, etc. (Portella, 2014). In a narrow sense, the 5D indicators, namely (population) density, (land-use) diversity, (land-use) design, traffic distance, and destination accessibility, as common built environment features, are often considered related to traffic risk (Ewing & Cervero, 2010). These characteristics are emphatically related to land use and roadway and road network, which directly affect traffic volume and speed (Huang, Wang, and Patton, 2018).

* Land-Use

Regarding land use, the land use diversity, specific land use types, and the distance to facilities corresponding to land use are related to traffic accident risks. Specifically, Wang and Kockelman (2013) introduced land use entropy as an index to measure land use balance(diversity) and found that more balanced land development had a slightly positive effect on reducing severe crashes. Similarly, Chen and Shen (2016) believe that higher land use diversity in the region can reduce the risk of cycling by reducing the travel distance, thereby reducing the collision frequency of vehicle-bicycle collisions. Ukkusuri et al. (2012) found that specific land use types, such as industrial and commercial, have a greater probability of pedestrian collisions. Proximity to specific services, such as schools, transport stops, etc., increased the risk of crashes (Pljakić, Jovanović and Matović, 2022; Mukoko & Pulugurtha, 2019). In addition, in the study of Wang et al. (2016), the land use intensity calculated based on the volume ratio and average building density is also found to be related to more pedestrian collision accidents.

* Roadway and Road Network

For road traffic, on the one hand, a number of empirical studies have shown that road shape, size, road service facilities and visual landscape are all related to road collision accidents (Gårder, 2004; Ukkusuri et al., 2012; Marshall & Garrick, 2012; Hanson, Noland & Brown, 2013). Ukkusuri et al. (2012) found that the probability of a pedestrian-vehicle collision increases with the number of lanes and road width. In turn, narrower lane designs with more pedestrians can improve road safety (Ewing & Dumbaugh, 2009). The geometry of road junctions is also considered to have a possible impact on collision accidents(Asgarzadeh et al., 2017). Collisions at non-orthogonal junctions are more likely to result in higher injury severity than at orthogonal junctions. In addition, multiple pedestrian infrastructures attached to roads, such as sidewalks, footpaths, marked crossings, overpasses and underpasses, are believed to reduce the risk of road accidents by reducing pedestrian contact with vehicular traffic and reducing vehicle speeds (Zegeer & Bushell , 2012; Stoker et al., 2015). Hanson, Noland and Brown (2013) tried to extract the visual features of road junctions based on the Google Street View Images, and estimated the relation between junctions' characteristics and frequency of injurious collisions. It is found that traffic islands, visual advertisements, bus stops and pedestrian crossings infrastructure are related to the increased pedestrian injuries in New York City.

On the other hand, the overall design characteristics of the road network are considered to regulate the speed and flow of vehicles on the road, and affect the risk of road accidents (Ewing and Dumbaugh, 2009). Most studies agree that dense, compact street networks improve safety. Marshall and Garrick (2012) used a negative binomial regression model to assess the effect of roadway and road network characteristics on crashes of different severity levels and found that a denser road network was associated with fewer crashes across all severity levels. Zhang et al. (2015) tested the association between various network structure variables and non-motor vehicle accidents, and believed that a denser road network design is safer. Nonetheless, this argument is questioned in recent empirical studies. Choi and Ewing (2021) found that denser and more connected network designs were associated with lower traffic volume but not significantly associated with crash rates.

In addition, the connectivity of roads in the network system was also found to be associated with accident risk. Some studies suggest that neighbourhoods with more connected streets are less safe regarding accident rates (Guo et al., 2017; Huang et al., 2018; Rifaat & Tay, 2009). Space syntax, a commonly used analysis theory and method for road network configuration (Hillier, 2007), has also been used to predict road traffic flow and explore the impact of road network design on accident risk. Guo et al. (2017) found that higher integration(closeness centrality) of specific road segments in a road network, is associated with more pedestrian-vehicle accidents. In a recent study, choice(betweeness centrality) was found to be associated with higher pedestrian mortality risk, although this relationship was non-linear and had local heterogeneity (Chang et al., 2022).

* Analysis units

Most of the above studies apply analysis units of large-scale geographical regions, administrative boundaries or statistical grids, and contribute to the overall understanding of the impact of built environment features. These explantary studies with high grainity are instructive for specifying macroscopic traffic planning and governance strategies (Ziakopoulos & Yannis, 2020), but it is often difficult to translate the knowledge gained into precise accident prevention strategies in driving scenarios or spatial locations.

Specifically, in studies that use geographic regions as the analysis unit, a factor that is difficult to observe is the precise location of the accident. This is related to spatial dependence and heterogeneity bias, which affects the performance of collision analysis (Xu & Huang, 2015). The same built environment factors may simultaneously have compound or diametrically opposite effects on traffic accidents in heterogeneous spatial units (Asadi et al., 2022). This increases the difficulty of proposing accurate and effective accident prevention measurements. Therefore, in addition to studying the built-environment causes of traffic accidents at a larger spatial granularity, an equally valuable question is, which built-environment factors may be associated with accident locations and traffic black spots? Research in this area is relatively limited.

## Identification and prediction of Traffic Black Spots

Generally, traffic black spots can be identified through historical accident records (Cui et al., 2022). Objects considered as traffic black spots can be road junctions, short road sections or mid-blocks (Kowtanapanich, Tanaboriboon, & Chadbunchachai, 2006). Common defining elements of traffic black spots may include the frequency of accidents in a specific spatial and temporal range, the severity of accidents, and the proportion of accident numbers to the traffic volume (Aziz & Ram, 2022). Furthermore, according to the differences in the spaital and social contexts, criteria for definning traffic black spots in regions can vary globally. For example, China defines urban road black spots as: road sections within 500 meters or junctions within 150 meters that have at least three casualty accidents every year (Yuan, Zeng & Shi, 2020); In Thailand, the black spot is the place where at least 3 serious accidents or 5 injury accidents have been recorded within 100m within three years (Tanprasert, T. et al., 2020).

Some studies do not adopt fixed threshold standards, but identify the clustering structure of collision accidents as traffic black spots. Modified clustering algorithm of K-Means (Zhang, Shu & Yan, 2019), Firefly (Yuan, Zeng & Shi, 2020) and DBSCAN(Ye et al., 2010) have been tested to improve the accuracy of identitfying traffic black spots from large accident datasets. However, in sparse or limited accident datasets the application of clustering methods is relatively limited.

With the advancement of information technology, methods for perceiving the urban environment and activities have been greatly enriched. Some studies seek to identify and predict traffic blackspots without relying on accident data, but through open data sources, such as street view images and built environment information from public map platforms. The popularity of various machine learning models also play an essential role in this trend. The visual information from street view image is believed to be useful for simulating and reflecting features of the built environment in the perception of drivers, cyclists or pedestrians(Ito and Biljecki, 2021; Biljecki and Ito, 2021). Some studies have attempted to predict street safety risks or the existence of traffic black spots based on street view elements and built environment information. For example, Tanprasert and his colleagues have explored a adistance-aware image segmentation model to extract visual information from street view objects and classify traffic black spots in Thailand based on a fully-connected neural network (Tanprasert et al., 2020). The model has shown an accuracy of 69.91% in classifying black and safe spots. Cai et al. (2022) extracted different street view objects from street view images and estimate the distance of object from the camera to simulate the perception of street view while driving. Street view elements combined with other built-environment variables were incorporated into three different tree-based ensemble models to predict the number of speeding accidents. The combination of street view information and the classification task of traffic black spots is proved to have sufficient potential.

# Case Study and Data

## Case Study Area

The research takes the London Borough of Camden as the case study area. Camden is one of the 12 boroughs located in Inner London (London Government Act 1963), and has always been an urban economic and cultural centre with active transport activities. Camden is a typical case study area for three reasons:

Firstly, Camden covers a wide variety of land uses, which shaped very different road environments. There is densely distributed cultural and commercial land in the south, the high streets and mixed-use areas of Camden Town and Kentish Town in the middle, and the residential areas and green spaces in Hampstead Heath in the north. The diversity in landuse enabled Camden to cover most road and road junction types in London.

Secondly, Camden is faced with significant pressure for traffic risk control. According to the 2021 census, the resident population of Camden is 210,100 (Office for National Statistics, 2022). Among London boroughs, Camden has a relatively higher accident casualty rate per capita (Transport for London, 2022, p.13). For several consecutive years during 2013-2021, Camden witnessed overall higher annual increases in the proportion of accident casualties than that of Inner and Greater London (ibid.).

Thirdly, the Camden authorities have successively introduced policies to create a safer traffic environment. The policies include speed limits on road sections near schools and on residential streets, adding flexible pedestrian and cycling safety facilities, and monitoring the risk level of road sections based on interactive maps (Safe Travel Camden, 2021). These policies generally show an emphasis on the optimisation of traffic facilities and the improvement of road safety based on precise geographic location.

|  |
| --- |
|  |
| **Figure 1.** Location of Camden Borough in the Greater London and the Selected Road Junctions |

## Data Collection

Considering that the COVID -19 may have a significant impact on long-term stability of traffic patterns (Transport for London, 2022), the research mainly uses the road accident data recorded in Camden before the COVID -19 as the research object. The research is mainly based on the 'Road Collision Attendants In Camden' (RCAC) dataset,which contains 4555 road collisions in Camden during 2015-2019 (London Borough of Camden, 2021). The dataset records the time, location, road type, speed limit, junction detail and other external environment details about the accidents and the severity class(fatal, serious, slight) and casualty number of accidents. Table 1 provides the definition of severity class by the Department of Transportation in the UK (GOV.UK, 2020).

Besides, the research also applies the road central line data from the Ordnance Survey (OS), the Point of Interest (POI) data from OpenStreetMap (OSM) and the administrative boundary data of Camden town. An image segmentation model is trained based on the Cambridge-driving Labelled Video Database(CamVid)， and street view images of road junctions are collected from Google Street View Static API.

|  |  |
| --- | --- |
| **Table 1.** The Definition of Accident Severity | |
| Severity | Definition |
| Fatal Accident | An accident in which at least one person is killed; other casualties (if any) may have serious or slight injuries. |
| Serious Accident | One in which at least one person is seriously injured, but no person is killed. |
| Slight Accident | One in which at least one person is slightly injured, but no person is killed or seriously injured. |

# Methods

## Research Framework

Figure 2 illustrates the research framework of this study. The research can be divided into three parts.

Firstly, according to the accident details included in the RCAC dataset, all the road junctions can be classified as safe, low-risk and high-risk. The classification is based on the spatial-temporal relationships between road accidents and junctions, and the accident severity.

Secondly, the research searches the built environment information around road junctions from a broad context and summarises the information into four dimensions. A series of analysis skills and models are applied in the process, including the Geopandas-based spatial data processing, the space syntax segment model for road network analysis, and the U-net deep learning model for semantic segmentation and street view elements extraction.

Finally, taking the junction risk levels as dependent variables and the four dimensions of built environment features as independent variables, multiple machine learning classifiers, such as Support Vector Machine (SVM), Random Forest (RF) and XGBOOST, are trained and compared to find the best classification model for accident risks prediction. Feature importance analysis is applied to understand the contributions of different built environmental features on model prediction.

|  |
| --- |
|  |
| **Figure 2.**The Research Framework of this Study |

## Measurement

### Road Junctions' Risk as Traffic Black Spots

Define road accidents within 30m from the junctions' geometric centre as accidents spatially related to junctions. According to the RCAC dataset, for each year during 2015-2019, about 70% of accidents happened near a road junction (Figure 3). The significantly high proportion indicates a higher risk of road junctions as traffic black points. Moreover, for each year the majority of the accidents are in the 'slight' type, and the 'serious' and 'fatal' types account for no more than 20% of the total accident types.

|  |  |
| --- | --- |
| 图表, 条形图  描述已自动生成 | 图表, 条形图  描述已自动生成 |
| **Figure 3.** Proportion of Accidents within and without the 30m Radius of Road Junctions | **Figure 4.** Proportion of Accidents in Different Severity Types |

Based on the count and severity of accidents happening around the junctions, 1877 road junctions within the Camden borough are divided into three risk levels (Figure 5). There are 1008 junctions with no collision accidents, which is classified as 'safe'; 445 junctions with no more than 2 slight accidents and no serious or fatal accident are classified as 'low-risk'; 424 junctions with more than 2 slight accidents, or with at least one serious or fatal accident, are classified as 'high-risk'.

|  |
| --- |
|  |
| **Figure 5.** Method Framework for Classifying Road Juncitons in Different Risk Types |

The figure below plots the spatial distribution of junctions with different risk levels. It is found that significantly more 'high-risk' junctions are located on the major roads in Camden. Most 'high-risk' and 'low-risk' junctions are concentrated in the middle and south of Camden, which are the local high streets and the historical centres for various urban functions. In contrast, 'safe' junctions are densely distributed on the minor roads in the north of Camden and many of them can be included as parts of local communities.

|  |
| --- |
|  |
| **Figure 6.** Spatial Mapping of Road Juncitons in Different Risk Types |

### Built Environment Features on the Road Junction

Four types of built environment features are selected for the study. According to the contexts where the features are collected, features can represent the junction's built environment information at a local or global scale. Min-Max scaling is applied to scale the numerical variables to 0-1.

**Table 2.** Classification Groups and Details of Different Built Environment Features

|  |  |  |  |
| --- | --- | --- | --- |
| **Group** | **Indicators** | **Spatial Scale** | **Measure** |
| **Land use and Activities** | POI diversity | Local | 0-1(Min-Max) |
| Number of culture POI in 50m |
| Number of food & drink POI in 50m |
| Number of hotel POI in 50m |
| Number of education POI in 50m |
| Number of shops & malls POI in 50m |
| Number of pub & bar POI in 50m |
| **Road Junction Structure and Facility** | Minimum distance to nearest junctions | Local | 0-1(Min-Max) |
| Count of roads connected |
| Maximum max speed of roads connected |
| Level of roads connected | Global | 0-All minor road, |
| 1-Major & minor road, |
| 2-All\_Major road |
| Presence of crossing | Local | 0-True, 1-False |
| Presence of traffic light |
| **Road Network Configuration** | Normalised Angular Choice R800m | Local | 0-1(Min-Max) |
| Normalised Angular Choice R3200m | Global |
| Normalised Angular Integration R800m | Local |
| Normalised Angular Integration R3200m | Global |
| **Street View** | Street view diversity | Local | 0-1(Min-Max) |
| Percent of building pixels |
| Percent of vegetation pixels |
| Percent of road pixels |
| Percent of sky pixels |
| Percent of sidewalk pixels |
| Percent of traffic sign pixels |

##### Activities Measurement

Points of Interest (POI) data provides direct knowledge of urban activities and land use information around road junctions at a local scale.

Through the open source map platform OpenStreetMap, POIs within a radius of 50m at each junction are retrieved and counted according to six categories, namely, food&drink, shop&mall, pub&bar, hotel, education, and culture. The number of POI in each category represents the vibrancy of the corresponding urban function around the junction. Based on the proportion of each POI type, a POI Diversity index (POID) is calculated as a measure of the complexity of urban activities around the junction.

|  |  |  |
| --- | --- | --- |
|  |  | (1) |
| Where, is the POI Diversity for n-th road junction; is the proportion of i-th type of POI in all the POIs around the n-th junction; when the count of POI around the n-th junction is equal or below 1, = 0. | | |

##### Junction Structure and Facility Measurement

The structure and facility characteristics of road junctions, are considered the important built environment features associated with accident risks. For structural features, the number of roads connected to the junction, determines the junction shape and basic traffic mode applied; the road level combinations on the junction, specifically, whether the junction is constituted of major roads, minor roads or a mix of them, affect the junction size and the importance of junction in the road system.

In terms of the facilities and functions related to road junctions, features are characterised as, whether there are traffic lights and crossing facilities at the junction, what is the maximum speed limit across roads connected to the junction, and what is the shortest distance for a junction to its nearest junction. In this study, it is hypothesised that sound traffic facilities can improve the safety of road junctions; while a loose speed limit and a short distance between adjacent junctions may complicate the traffic environment, thereby increasing the risk of accidents.

##### Road Network Measurement

Based on the space syntax theory, the multi-scale configuration characteristics of urban road networks can help reflect and describe the potential distribution of urban activities and the natural movement pattern of individuals in urban space (Hillier and Hanson, 1984; Hillier, 2007). A series of network metrics is developed to quantify the interaction between spatial configuration and the activity and movement. Among them, the Normalised Angular Integration (NAIN) is used to reflect the attractiveness of different road segments to urban activities at a specific spatial scale; the Normalised Angular Choice(NACH) reflects the potential distribution characteristics of traffic flow on different road segments in the road network (Hillier, Yang and Turner, 2012). The explanation of the calculation methods of NAIN and NACH can be found in the appendix. The research attempts to calculate the NAIN and NACH of each road segment in the Camden road network with 800m and 3600m as the analysis radius. The radius of 800m and 3600m correspond to the local and global scales around the junction, respectively. On this basis, the arithmetic mean of NAIN and NACH of the connected road segments is summarised for each junction in Camden, to reflect the potential attractiveness of road junctions to urban activities and traffic volume in the network.

##### Street View Measurement

Street View images of road junctions provide intuitive visual information about junctions' built environment characteristics at a local scale. Based on the geometric centre of 1877 road junctions, street view images are collected from Google Street View Static API in four equally spaced directions of 0°, 90°, 180°, and 270°.

Using Python's deep learning library, FastAI (Howard & Gugger, 2020), and the open-sourced computer vision dataset, CamVid (Brostow, Fauqueur and Cipolla, 2008), an image segmentation model is trained to identify and calculate the proportion of different street view elements for each street view image.

Six types of static street view labels are selected as the reflection of built environment features of junctions, including building (building, fence, wall), road (road, parking block, road shoulder), vegetation (tree and other vegetation types), traffic sign (traffic light, Sign & Symbol), sky and sidewalk. Some dynamic street view elements, such as vehicles and pedestrians, are considered less frequent in street view images and are more sensitive to image collection methods, which are not suitable for characterising the built environment of junctions (Kim, 2021).

For each junction, the mean proportion of each street view elements across the four images was calculated. On this basis, we tried to calculate the Street View Diversity (SVD) of each junction with the help of the Shannon Weaver Diversity index (SWD). As a comprehensive measure, SVD characterises the complexity and disorder of street view elements at the junction scale.

|  |  |  |
| --- | --- | --- |
|  |  | (1) |
| Where, is the Street View Diversity for n-th road junction; is the proportion of i-th type of selected street view features in all the street view features in the n-th junction; | | |

## Machine Learning Classification

Taking the four groups of built environment features of junctions as independent variables, and junctions'accident risk levels as dependent variables, multiple machine learning classifiers are fitted to identify and predict junctions' potential risk as traffic black spots.

The original data are randomly split into training and test datasets in the weight of 9:1. Considering that the proportion of safe junctions is much higher than that of low-risk and high-risk junctions, the Synthetic Minority Oversampling Technique (SMOTE) is applied to rebalance the training dataset (Chawla et al., 2002).

The study first compared the performance of six initial machine learning classifiers in junction risk classification tasks. Four groups of built environment features, in group and collectively, are input into the classifiers as independent variables. All of the classifiers are with default hyperparameter settings. The classifiers involved in the comparison are listed in Table 1. Among them, random forest classifiers are tested to have higher overall accuracy and lower log loss. The random forest is also considered a more explainable model for multi-object classification tasks (Rodriguez-Galiano et al., 2012). Therefore, the random forest classifier is selected as the core model for further parameter tuning and cross-validation.

**Table 3.** Performance Comparison for Different Machine Learning Classifiers Trained on All Built Environment Features and with Default Parameters

| **Classifier** | **Accuracy** | **Log Loss** |
| --- | --- | --- |
| KNeighborsClassifier | 62.765957 | 6.837019 |
| MLPClassifier | 65.957447 | 0.732221 |
| SVC | 69.148936 | 0.788977 |
| NuSVC | 67.553191 | 0.786658 |
| DecisionTreeClassifier | 57.446809 | 15.337725 |
| RandomForestClassifier | 68.085106 | 0.742107 |
| XGBClassifier | 64.893617 | 0.892583 |
| AdaBoostClassifier | 66.489362 | 1.064485 |
| GradientBoostingClassifier | 65.957447 | 0.759853 |
| GaussianNB | 67.021277 | 2.447175 |
| LinearDiscriminantAnalysis | 64.361702 | 0.757298 |
| QuadraticDiscriminantAnalysis | 59.042553 | 3.774714 |

The study used all four types of built-in environment variables to fit the best classifier, and applied the GridSearchCV method to tune the model based on different parameter combinations, with accuracy as scoring. For each set of parameter combinations, k-fold method is applied to spilt the training data set into a training set and a validation set for cross-validation. Finally, the study attempts to use the permutation feature importance method to identify specific built-environment variables with higher contribution in identifying the junction risk (Parr & Turgutlu, 2018). Considering the randomness included in the generation process of feature importance scores, the feature importance algorithm is run for 100 times with same parameters to calculate the mean scores as the general ranking of feature importance.

# Results

## Traffic Black Spots Classification

### Random Forst Classifier Performance

The accuracy performance of random forest classifiers trained based on the 4 groups of built environment features, in group and together, are shown in Figure 7. In general, the model trained based on all the feature groups has gained the highest accuracy score of 0.697 on the test datasets. Besides, some feature groups have shown a significantly stronger fit than other feature groups in classifying the junction's risk levels. Specifically, models fitted with multi-scale road network configuration measures gained a better performance in classifying the risk level of accidents, with an accuracy score of 0.644 on the test data set. In contrast, models fitted separately with the land use, street view and junction function and faciltiy characteristics have accuracy scores lower than 0.6. The POI-based model is under-fitted.



**Figure 7.** Performance Comparison for Random Forest Classifiers Trained on Different Feauture Groups

The study selected the random forest classifier fitted with all the features for model tuning and cross-validation. The GridSearchCV method from the Sklearn library is applied to search for the best hyperparameter combination for the classifier from a preset searching grid of multiple 'max\_depth' and 'n\_estimator' values. K-fold method is applied in cross-validation, and the fold number is set to 5.

**Table 4.** The Parameter Searching Grid and the Best Parameters Gained

|  |  |  |
| --- | --- | --- |
| Parameter Type | Searching Grid | Best Parameters |
| max\_depth | 10, 60, 110, 160, 210, none | 60 |
| n\_estimators | 50, 100, 150, 200, 250, 300, 350, 400, 450, 500 | 500 |

According to the analysis, the best hyperparameters couple are 60 for 'max\_depth' and 500 for 'n\_estimators'. The performance of the corresponding classifier on the test dataset is shown in Table. Overall, the classifier has an accuracy score of 0.71 and presents strong abilities in identifying safe and high-risk junctions. Specifically, the model shows the best performance in classifying safe junctions, where the f1-scores is 0.84, presenting a good balance in both idntifying and recall all the safe junctions. However, in classifying junctions with detailed risk levels, the performance of the model is slightly weaker. Model gains relatively high f1-scores in identifying high-risk junctions, but poor scores in classifying low-risk junctions.

**Table 5.** The Parameter Searching Grid and the Best Parameters Gained

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **precision** | **recall** | **f1-score** | **support** |
| **0** | 0.80 | 0.89 | 0.84 | 100 |
| **1** | 0.34 | 0.26 | 0.29 | 39 |
| **2** | 0.73 | 0.71 | 0.72 | 49 |
| **Accuracy** |  | | 0.71 | 188 |
| **Macro avg** | 0.63 | 0.62 | 0.62 | 188 |
| **Weighted avg** | 0.69 | 0.71 | 0.70 | 188 |

### Feature Importance

The box plots generated based on the importance scores for each feature are shown in the Figure 7. According to the average value of the importance score, the ranking of the feature importance of different groups is roughly as follows: **Road Network Configuration** >**Street View** >**Junction Structure and Facility**> **Land Use and Activities**.

Specifically，three configuration features, namely, NAINr800, NACHr3200 , and NAIN r3200 ranked top 3 as the most important features in the random forest model. The three features are closely related to the overall structure of the road network. Among them, the NAIN r800 and NAIN r3200 represent the local and global measurement of the spatial attractions of junctions; while the NACH r3200 represents the global measuremnet of the importance of junctions as potential traffic hub. It is noticed that the variance of their importance score across different tests is also large. The three configuration features are followed by a more direct feature representing the global location and structure of junctions, that is whether the junction is connected only by minor roads. The local-scale measured choice values ranked 5th.

Beyond the 5 features, generally, the junctions' SVI features are more important than the structure and facility features of junctions. Typically, sky, building and sidewalks are the most important features in the SVI feature group. The minmum distance between junctions, and the existense of crossing and traffic light also ranked higher in features related to the structure and facility of junctions. The complexity of street view features, and almost all the POI related features are not so important in the classification model. Though findings from feature importance above, feature importance only refers to how features are useful when building a specific model and should not be interpreted as direct dependencies between predictors and targets.

图表

描述已自动生成

**Figure 8.** Boxplot of the Feature Importance Scores

# Discussion

**Global Features vs Local Features**

In this study, it is found that variables capturing more global information about road and urban systems have a stronger predictive capability for junction risk. Specifically, among the default random forest classifiers based on single variable group, the model based on road network configuration features, which represent the multi-scale spatial connections performed the best. In the tuned model based on all variables, the features of NAINr800, NAINr3200 and NACHr3200 , and the grades of roads participating in the formation of junctions, are genrally more important than the features only representing the local characteristics of the built environment. The typical local features gained poor performance include the number of different POIs at junciotns, and trees, walls and other visual elements in the juncitons’ street view. The finding suggests that traffic risk at road junctions should be considered more as a systemic problem to be analysed and managed. The role of junctions in the road network system is possibly the most important factor deciding the safety of a jucniton.

Compared to the previous study (Guo et al., 2017; Chang et al., 2022), this study proves again that both the integration and choice measures of the road network are associated with traffic accident risks. Moreover, the scale difference of these network configuration matters in the predictive performance of risk. The choice measured within a larger analysis radius and the integration measured within a smaller analysis radius, gain better performance in the fitted model.

**Sense vs Semantics**

In this study, the classifier fitted with street view information has relatively weaker performance than the model with configuration features, but has stronger performance than models based on abstract geometry or semantics features, such as junction structure and facilitis and land use and activities.

On the one hand, different from the better profermance in the previous study of Tanprasert et al (2020).

the poor performance of street view images in this study, may be due to the research: 1. Only junctions are used as the research object, and there is a convergence of street view features. 2. There are differences in the definition and classification methods of traffic black spots.

On the other hand, in the comprehensive model, road, roadshoulder, building, sky, etc. in the street view information still have relatively important contributions to the classification of junction risks, and their contribution is higher than the indicators obtained based on traditional built environment measurement methods. This shows that the perception of urban space via Street View images is a reliable and efficient method.

# Conclusion

## Findings

Here is a summary of the findings in this study. In different machine learning classifiers with default parameters, the random forest classifier is tested as the best performing classifers towards junction risk classification task based on the built environment features. For RF models fitted with single group of built environment features, the model fitted with road network configuration variables performed best in prediction. Furthermore, the performance of model with all variables has a small but significant improvement in accuracy compared to the best model fitted with a single feature group. This indicates that other environmental information and factors also contribute to the improvement of model progress.

The tuned model random forest model fitted with all built environment features has an accuracy of 0.71 on the test data set, and has a good classification performance for safe and high-risk junctions, but weak classification ability for low-risk junctions. Whether the junction is only based on the minor road, and the multi-scale NAIN and NACH features of the junction, occupy the first place in the ranking of feature importance. The importance ranking among other feature group is generally the street view of the junction, the structure and facility features of the junction, and the landuse activity feature around the junction. Considering that the important goal of building a predictive model in this study is to identify potential high-risk junctions in current urban management and future urban road design, reduce and prevent casualties, and despite its flaws, the model as a whole is effective.

## Innovation

The study applies space syntax method for road network configuration analysi and is the first combination of space syntax and machine learning for risk prediction and classification for road junctions. To some extent，space syntax provides intuitive knowledge of the status and role of individual juncitons with respect to street networks and junction systems. By combining the global perception with space syntax and the local perception from street view images, the study has the opportunity to gain a more comprehensive understanding of how urban built environment systems composed of similar elements and unified rules, will interact with traffic accident risks. This research also provides a reference for the management and maintenance of current road junctions, as well as the design of future road networks and junctions.

## Limitations

Although this study set three junction risk levels, only safe junctions and high-risk junctions can be effectively classified. This is mainly due to the fact that current accident data is sparse within the Camden Borough area. The model learns the characteristics of low-risk junctions insufficiently.

# Acknowledgements

# References

Asadi, M. *et al.* (2022) ‘A comprehensive analysis of the relationships between the built environment and traffic safety in the Dutch urban areas’, *Accident Analysis & Prevention*, 172, p. 106683. Available at: <https://doi.org/10.1016/j.aap.2022.106683>.

Asgarzadeh, M. *et al.* (2017) 'The role of intersection and street design on severity of bicycle-motor vehicle crashes', *Injury Prevention*, 23(3), pp. 179–185. Available at: <https://doi.org/10.1136/injuryprev-2016-042045>.

Aziz, S. and Ram, S. (2022) ‘A Meta-analysis of the methodologies practiced worldwide for the identification of Road Accident Black Spots’, *Transportation Research Procedia*, 62, pp. 790–797. Available at: <https://doi.org/10.1016/j.trpro.2022.02.098>.

Biljecki, F. and Ito, K. (2021) ‘Street view imagery in urban analytics and GIS: A review’, *Landscape and Urban Planning*, 215, p. 104217. Available at: <https://doi.org/10.1016/j.landurbplan.2021.104217>.

Brostow, G.J. *et al.* (2008) ‘Segmentation and Recognition Using Structure from Motion Point Clouds’, in D. Forsyth, P. Torr, and A. Zisserman (eds) *Computer Vision – ECCV 2008*. Berlin, Heidelberg: Springer (Lecture Notes in Computer Science), pp. 44–57. Available at: <https://doi.org/10.1007/978-3-540-88682-2_5>.

Chang, I. *et al.* (2022) ‘Predicting effects of built environment on fatal pedestrian accidents at location-specific level: Application of XGBoost and SHAP’, *Accident Analysis & Prevention*, 166, p. 106545. Available at: <https://doi.org/10.1016/j.aap.2021.106545>.

Chawla, N.V. *et al.* (2002) ‘SMOTE: Synthetic Minority Over-sampling Technique’, *Journal of Artificial Intelligence Research*, 16, pp. 321–357. Available at: <https://doi.org/10.1613/jair.953>.

Chen, P. and Shen, Q. (2016) ‘Built environment effects on cyclist injury severity in automobile-involved bicycle crashes’, *Accident Analysis & Prevention*, 86, pp. 239–246. Available at: <https://doi.org/10.1016/j.aap.2015.11.002>.

Choi, D. and Ewing, R. (2021) ‘Effect of street network design on traffic congestion and traffic safety’, *Journal of Transport Geography*, 96, p. 103200. Available at: <https://doi.org/10.1016/j.jtrangeo.2021.103200>.

Cui, H. *et al.* (2022) ‘Identifying accident black spots based on the accident spacing distribution’, *Journal of Traffic and Transportation Engineering (English Edition)*, 9(6), pp. 1017–1026. Available at: <https://doi.org/10.1016/j.jtte.2021.02.006>.

Dereli, M.A. and Erdogan, S. (2017) ‘A new model for determining the traffic accident black spots using GIS-aided spatial statistical methods’, *Transportation Research Part A: Policy and Practice*, 103, pp. 106–117. Available at: <https://doi.org/10.1016/j.tra.2017.05.031>.

Ewing, R. and Dumbaugh, E. (2009) ‘The Built Environment and Traffic Safety: A Review of Empirical Evidence’, *Journal of Planning Literature*, 23(4), pp. 347–367. Available at: <https://doi.org/10.1177/0885412209335553>.

Ewing, R. and Cervero, R. (2010) ‘Travel and the Built Environment: A Meta-Analysis’, *Journal of the American Planning Association*, 76(3), pp. 265–294. Available at: <https://doi.org/10.1080/01944361003766766>.

Gårder, P.E. (2004) ‘The impact of speed and other variables on pedestrian safety in Maine’, *Accident Analysis & Prevention*, 36(4), pp. 533–542. Available at: <https://doi.org/10.1016/S0001-4575(03)00059-9>.

Gitelman, V. *et al.* (2017) ‘Exploring Safety Impacts of Pedestrian Crossing Configurations at Signalized Junctions on Urban Roads with Public Transport Routes’, *Transportation Research Procedia*, 25, pp. 2044–2060. Available at: <https://doi.org/10.1016/j.trpro.2017.05.399>.

GOV.UK (2020).*Reported road casualties in Great Britain: notes, definitions, symbols and conventions*. Available at: <https://www.gov.uk/government/publications/road-accidents-and-safety-statistics-notes-and-definitions/reported-road-casualties-in-great-britain-notes-definitions-symbols-and-conventions> (Accessed: 28 January 2023).

Guo, Q. *et al.* (2017) ‘The effect of road network patterns on pedestrian safety: A zone-based Bayesian spatial modeling approach’, *Accident Analysis & Prevention*, 99, pp. 114–124. Available at: <https://doi.org/10.1016/j.aap.2016.11.002>.

Hanson, C.S., Noland, R.B. and Brown, C. (2013) ‘The severity of pedestrian crashes: an analysis using Google Street View imagery’, *Journal of Transport Geography*, 33, pp. 42–53. Available at: <https://doi.org/10.1016/j.jtrangeo.2013.09.002>.

Hillier, B. and Hanson, J. (1984) *The Social Logic of Space*. Cambridge: Cambridge University Press. Available at: <https://doi.org/10.1017/CBO9780511597237>.

Hillier, B. (2007) *Space is the machine: a configurational theory of architecture*. Space Syntax.

Hillier, B., Yang, T. and Turner, A. (2012) 'Advancing DepthMap to advance our understanding of cities: comparing streets and cities and streets with cities', in *8th International Space Syntax Symposium. Pontificia Universidad Catolica de Chile: Santiago, Chile.* Available at: <https://www.mysciencework.com/publication/show/advancing-depthmap-advance-our-understanding-cities-comparing-streets-cities-streets-cities-d1cc615a> (Accessed: 5 February 2023).

Howard, J. and Gugger, S. (2020) ‘Fastai: A Layered API for Deep Learning’, *Information*, 11(2), p. 108. Available at: <https://doi.org/10.3390/info11020108>.

Huang, Y., Wang, X. and Patton, D. (2018) ‘Examining spatial relationships between crashes and the built environment: A geographically weighted regression approach’, *Journal of Transport Geography*, 69, pp. 221–233. Available at: <https://doi.org/10.1016/j.jtrangeo.2018.04.027>.

Ito, K. and Biljecki, F. (2021) ‘Assessing bikeability with street view imagery and computer vision’, *Transportation Research Part C: Emerging Technologies*, 132, p. 103371. Available at: <https://doi.org/10.1016/j.trc.2021.103371>.

Kim, J.H. *et al.* (2021) ‘Decoding urban landscapes: Google street view and measurement sensitivity’, *Computers, Environment and Urban Systems*, 88, p. 101626. Available at: <https://doi.org/10.1016/j.compenvurbsys.2021.101626>.

Kowtanapanich, W., Tanaboriboon, Y. and Chadbunchachai, W. (2006) ‘APPLYING PUBLIC PARTICIPATION APPROACH TO BLACK SPOT IDENTIFICATION PROCESS: – A Case Study in Thailand –’, *IATSS Research*, 30(1), pp. 73–85. Available at: <https://doi.org/10.1016/S0386-1112(14)60158-8>.

Kullgren, A. *et al.* (2019) ‘The potential of vehicle and road infrastructure interventions in fatal bicyclist accidents on Swedish roads—What can in-depth studies tell us?’, *Traffic Injury Prevention*, 20(sup1), pp. S7–S12. Available at: <https://doi.org/10.1080/15389588.2019.1610171>.

Leslie, W.I. (1977) ‘IDENTIFICATION OF ROAD ACCIDENT BLACK SPOTS’, in. *Compendium of Technical Papers of the 47th Annual Meeting of the Institute of Transportation Engineers and the Fourth World Transportation Engineers Conference, Mexico City, October 2-6, 1977.* Available at: <https://trid.trb.org/view/81117> (Accessed: 27 January 2023).

London Government Act 1963, c. 33. Available at: <https://www.legislation.gov.uk/ukpga/1963/33> (Accessed: 28 January 2023).

Marshall, W.E. and Garrick, N.W. (2011) ‘Does street network design affect traffic safety?’, *Accident Analysis & Prevention*, 43(3), pp. 769–781. Available at: <https://doi.org/10.1016/j.aap.2010.10.024>.

Mukoko, K.K. and Pulugurtha, S.S. (2020) ‘Examining the influence of network, land use, and demographic characteristics to estimate the number of bicycle-vehicle crashes on urban roads’, *IATSS Research*, 44(1), pp. 8–16. Available at: <https://doi.org/10.1016/j.iatssr.2019.04.001>.

Wang, Y. and Kockelman, K.M. (2013) ‘A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods’, *Accident Analysis & Prevention*, 60, pp. 71–84. Available at: <https://doi.org/10.1016/j.aap.2013.07.030>.

World Health Organisation. (2018) *Global status report on road safety.* Available at: <https://www.who.int/publications-detail-redirect/9789241565684> (Accessed: 10 February 2023).

Office for National Statistics (2022). *Population and household estimates, England and Wales: Census 2021.* Available at: <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationandhouseholdestimatesenglandandwalescensus2021> (Accessed: 28 January 2023).

Osama, A. and Sayed, T. (2017) ‘Evaluating the Impact of Socioeconomics, Land Use, Built Environment, and Road Facility on Cyclist Safety’, *Transportation Research Record*, 2659(1), pp. 33–42. Available at: <https://doi.org/10.3141/2659-04>.

Parr, T. and Turgutlu, K. (2023) ‘Feature importances for scikit-learn machine learning models’. Available at: <https://github.com/parrt/random-forest-importances> (Accessed: 5 February 2023).

Pljakić, M., Jovanović, D. and Matović, B. (2022) 'The influence of traffic-infrastructure factors on pedestrian accidents at the macro-level: The geographically weighted regression approach', *Journal of Safety Research*, 83, pp. 248–259. Available at: <https://doi.org/10.1016/j.jsr.2022.08.021>.

Portella, A.A. (2014) ‘Built Environment’, in A.C. Michalos (ed.) *Encyclopedia of Quality of Life and Well-Being Research*. Dordrecht: Springer Netherlands, pp. 454–461. Available at: <https://doi.org/10.1007/978-94-007-0753-5_240>.

Ren, H. *et al.* (2017) ‘A Deep Learning Approach to the Prediction of Short-term Traffic Accident Risk’.

Rifaat, S.M. and Tay, R. (2009) ‘Effects of Street Patterns on Injury Risks in Two-Vehicle Crashes’, *Transportation Research Record*, 2102(1), pp. 61–67. Available at: <https://doi.org/10.3141/2102-08>.

Rodriguez-Galiano, V.F. *et al.* (2012) ‘An assessment of the effectiveness of a random forest classifier for land-cover classification’, *ISPRS Journal of Photogrammetry and Remote Sensing*, 67, pp. 93–104. Available at: <https://doi.org/10.1016/j.isprsjprs.2011.11.002>.

Safe Travel Camden (2021) *About Safe Travel Camden*. Available at: <https://safetravelcamden.commonplace.is/about> (Accessed: 28 January 2023).

Siddiqui, C., Abdel-Aty, M. and Choi, K. (2012) ‘Macroscopic spatial analysis of pedestrian and bicycle crashes’, *Accident Analysis & Prevention*, 45, pp. 382–391. Available at: <https://doi.org/10.1016/j.aap.2011.08.003>.

Stoker, P. *et al.* (2015) 'Pedestrian Safety and the Built Environment: A Review of the Risk Factors', *Journal of Planning Literature*, 30(4), pp. 377–392. Available at: <https://doi.org/10.1177/0885412215595438>.

Tanprasert, T. *et al.* (2020) ‘Recognizing Traffic Black Spots From Street View Images Using Environment-Aware Image Processing and Neural Network’, *IEEE Access*, 8, pp. 121469–121478. Available at: <https://doi.org/10.1109/ACCESS.2020.3006493>.

Transport for London (2022). Casualties in Greater London during 2021 . Available at: <https://content.tfl.gov.uk/casualties-in-greater-london-2021.pdf> (Accessed: 28 January 2023).

Ukkusuri, S. *et al.* (2012) ‘The role of built environment on pedestrian crash frequency’, *Safety Science*, 50(4), pp. 1141–1151. Available at: <https://doi.org/10.1016/j.ssci.2011.09.012>.

Wang, Y. and Kockelman, K.M. (2013) ‘A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods’, *Accident Analysis & Prevention*, 60, pp. 71–84. Available at: <https://doi.org/10.1016/j.aap.2013.07.030>.

Wang, X. *et al.* (2016) ‘Macro-level safety analysis of pedestrian crashes in Shanghai, China’, *Accident Analysis & Prevention*, 96, pp. 12–21. Available at: <https://doi.org/10.1016/j.aap.2016.07.028>.

Xu, P. and Huang, H. (2015) ‘Modeling crash spatial heterogeneity: Random parameter versus geographically weighting’, *Accident Analysis & Prevention*, 75, pp. 16–25. Available at: <https://doi.org/10.1016/j.aap.2014.10.020>.

Ye, J. *et al.* (2010) ‘Research and implement of traffic accident analysis system based on accident black spot’, in *2010 5th International Conference on Computer Science & Education*. *2010 5th International Conference on Computer Science & Education*, pp. 1805–1809. Available at: <https://doi.org/10.1109/ICCSE.2010.5593800>.

Yuan, T., Zeng, X. and Shi, T. (2020) ‘Identifying Urban Road Black Spots with a Novel Method Based on the Firefly Clustering Algorithm and a Geographic Information System’, *Sustainability*, 12(5), p. 2091. Available at: <https://doi.org/10.3390/su12052091>.

Zegeer, C.V. and Bushell, M. (2012) 'Pedestrian crash trends and potential countermeasures from around the world', *Accident Analysis & Prevention*, 44(1), pp. 3–11. Available at: <https://doi.org/10.1016/j.aap.2010.12.007>.

Zhang, C., Shu, Y. and Yan, L. (2019) ‘A Novel Identification Model for Road Traffic Accident Black Spots: A Case Study in Ningbo, China’, *IEEE Access*, 7, pp. 140197–140205. Available at: <https://doi.org/10.1109/ACCESS.2019.2942647>.

Zhang, Y. *et al.* (2015) ‘Investigating the associations between road network structure and non-motorist accidents’, *Journal of Transport Geography*, 42, pp. 34–47. Available at: <https://doi.org/10.1016/j.jtrangeo.2014.10.010>.

Ziakopoulos, A. and Yannis, G. (2020) ‘A review of spatial approaches in road safety’, *Accident Analysis & Prevention*, 135, p. 105323. Available at: <https://doi.org/10.1016/j.aap.2019.105323>.

Overall comments:

1. The current classification results are good to me, and we don't need to change the datasets
2. The tuned model looks good, as we use kfold validation and train-test split to tune the model. No need to compare the performance of the tuned and default model.
3. This paper is quite long, especially in Section 6. Some maps in Section 6 can be moved to appendix or removed. the exploratory analysis of the input variables can be shortened. (we can suspend the writing of this part and first focus on results and discussion)
4. Some vars & acronyms in SHAP plot are not described, like 'all minor\_1', 'all major\_3'.
5. Be cautious about discussing 'when traffic volume data on the junction is absent', as we didn't compare the model performance with and without traffic volume data. Moreover, road configuration is not a replacement for traffic volume data. Moreover, TfL has the traffic volume data. I would rather not mention 'traffic volume data' in this paper.

Next step of writing:

1. In writing 6.2.2 and Section 7, you can use simple sentences to communicate the key points (rather than complicated sentences). I added many comments to this part.
2. We don't want to make the discussion very complicated, so please focus on the key points.

# Appdendix

# Traffic Accident Conditions in Camden

|  |  |  |  |
| --- | --- | --- | --- |
| **Local Area** | **Casualties** | **Population** | **Ratio of casuality** |
| **Greater london** | 26,406 | 8,800,000 | 0.30% |
| **Inter london** | 10907 | 2,789,100 | 0.39% |
| Camden | 767 | 210,100 | 0.37% |
| City of London | 152 | 8,600 | 1.77% |
| Hackney | 970 | 259,200 | 0.37% |
| Hammersmith and Fulham | 687 | 183,200 | 0.38% |
| Islington | 724 | 216,600 | 0.33% |
| Kensington and Chelsea | 642 | 143,400 | 0.45% |
| Lambeth | 1,343 | 317,600 | 0.42% |
| Lewisham | 843 | 300,600 | 0.28% |
| Southwark | 1,097 | 307,700 | 0.36% |
| Tower Hamlets | 1,276 | 310,300 | 0.41% |
| Wandsworth | 1,135 | 327,500 | 0.35% |
| Westminster | 1,271 | 204,300 | 0.62% |
| **Outer london** | 15,499 | 6,010,900 | 0.26% |
| Barking and Dagenham | 615 | 218,900 | 0.28% |
| Barnet | 1,014 | 389,300 | 0.26% |
| Bexley | 506 | 246,500 | 0.21% |
| Brent | 997 | 339,800 | 0.29% |
| Bromley | 741 | 330,000 | 0.22% |
| Croydon | 1,112 | 390,800 | 0.28% |
| Ealing | 1,065 | 367,100 | 0.29% |
| Enfield | 1,120 | 330,000 | 0.34% |
| Greenwich | 807 | 289,100 | 0.28% |
| Haringey | 683 | 264,200 | 0.26% |
| Harrow | 421 | 261,300 | 0.16% |
| Havering | 683 | 262,000 | 0.26% |
| Hillingdon | 708 | 305,900 | 0.23% |
| Hounslow | 771 | 288,200 | 0.27% |
| Kingston upon Thames | 358 | 168,000 | 0.21% |
| Merton | 508 | 215,200 | 0.24% |
| Newham | 1,032 | 351,100 | 0.29% |
| Redbridge | 828 | 310,300 | 0.27% |
| Richmond upon Thames | 416 | 195,200 | 0.21% |
| Sutton | 435 | 209,600 | 0.21% |
| Waltham Forest | 679 | 278,400 | 0.24% |

![图表, 瀑布图

描述已自动生成]()

# Built Environment Data Sources

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Group** | **Indicators** | **Data Source** | **Spatial Scale** | **Measure Scale** |
| **Junction Structure and Facility** | Minimum distance to nearest junctions | OS | Local | 0-1(Min-Max) |
| Count of roads connected |
| Maximum max speed of roads connected | OSM |
| Level of roads connected | OS | Global | 0-ALL minor road,  1-Major&minor road,  2-All\_Major road |
| Presence of crossing | OSM | Local | 0-True, 1-False |
| Presence of traffic light |
| **Street View** | Street view complexity | SVI | Local | 0-1(Min-Max) |
| Perception of building pixels |
| Percent of vegetation pixels |
| Percent of road pixels |
| Percent of sky pixels |
| Percent of sidewalk pixels |
| Percent of traffic sign pixels |
| **Comercial and Public Activities** | Number of cultural POI in 50m | OSM | Local | 0-1(Min-Max) |
| Number of food & drink POI in 50m |
| Number of hotels in 50m |
| Number of schools in 50m |
| Number of shops & malls in 50m |
| Number of pub & bar POI in 50m |
| **Road Network Configuration** | Normalised Angular Choice R800m | OS | Local | 0-1(Min-Max) |
| Normalised Angular Choice R3200m | Global |
| Normalised Angular Integration R800m | Local |
| Normalised Angular Integration R3200m | Global |

# Space Syntax Measurements

Space syntax is a set of architectural and urban analysis theories and methodologies rooted in graph theory. At the urban scale, Segment analysis based on the road network is a commonly used analysis method, where the road system as a whole is abstracted into an undirected graph, road segments are represented as nodes in the graph, and junctions are represented as edges between nodes. By controlling the analysis radius (distance threshold) in real space, the number of nodes covered in the graph can be adjusted, thereby controlling the scale and complexity of the graph. Normalised Angular Integration (NAIN) and Normalised Angular Choice (NACH) are classic measurements in segment analysis.

NAIN can be regarded as a development of the closeness centrality measure in graph theory for the spatial network. Mathematically, the closeness centrality can be illustrated as the quotient of the total number of nodes in the graph and the sum of the distances from a specific node i in the graph to any other node in the shortest path (that is, the reciprocal of the average depth of node i in the graph). The distance calculation of the depth is based on the number of edges contained in the path and the weight of the edges.

Differently, in segment analysis the graph is based on the characteristics of the real road network, and defines the weight of the edge as the angular deviation between roads. This method simulates a natural way-finding mode, where to search for the shortest path between two points, based on the current road segment, always find the next road segment with a smaller angle deflection. Usually, the straightest path is to be found. Modifying the formula parameters of closeness centrality, NAIN is characterised as:

|  |  |  |
| --- | --- | --- |
|  |  | (2) |
| represents the Normalised Angular Integration for the node i in the graph;represents the distance of the shortest path between node i and any node j in the graph, where the distance is measured by the change of angle between each two connected edges in the graph; *n* represents the number of nodes in the graph. | | |

Similarly, NACH was born out of the betweenness centrality in graph theory, which represents the possibility of node i being included in the shortest path between any two nodes in the graph. The basic betweenness centrality calculation is shown in formula (3). On this basis, NACH uses the total depth of the graph to standardise the betweenness. See formula (4).

|  |  |  |
| --- | --- | --- |
|  |  | (3) |
| represents the betweenness of node i; represents the times i lies on the shortest path from any node s to any node t; represents the total number of path from s to t. | | |

|  |  |  |
| --- | --- | --- |
|  |  | (4) |
|  | | |

represents the Normalised Angular Choice of the node i in the graph; represents the betweenness of node i; represents the distance of the shortest path between node i and any node j in the graph; The distance is measured by the change of angle between each two connected edges in the graph.

# Accident Count Distrbution by junctions

Figure 4 then plots all the road junctions and highlights their difference in accident count. The accident count follows an exponential distribution, where most of the junctions have no accident. Besides, it is found that most of the accidents happened on the junctions near local commercial centres or high streets in the southern area of Camden, which indicates a possible relation between commercial activities and accidents.

|  |
| --- |
|  |
| Figure 4 |

# Built Environment Features around Junctions

This section first presents the spatial distribution of different types of built environment features at road junctions within Camden. On this basis, for junctions with various risk levels, the one-way analysis of variance (ANOVA) method was applied to compare whether there is a significant difference between the built environment features.

##### **Junction Structure and function**

This section demonstrates the differences in the structural and functional characteristics of junctions within the Camden borough. In terms of structural features, most of the junctions are 4-way junctions, while 5-way junctions also occupy a considerable proportion. The number of 3-way and more than 5 ways junctions are limited. Most junctions are minor-minor junctions and major-minor mixed junctions, due to a low density of major roads in Camden borough.

In terms of functional characteristics, the crossings are more common junction facilities than traffic lights across Camden borough. Significantly more crossings and traffic lights are located at road junctions in active urban areas in central and southern Camden. The overall maximum speed limits of the roads connected to junctions range from 30km/h to 45km/h, and are subject to the spatial distribution of specific major roads. The distribution of the shortest distance from each junction to the nearby junctions is not spatially significant.

|  |  |  |  |
| --- | --- | --- | --- |
| 图表, 散点图  描述已自动生成 | | | |
| Minimum distance to nearest junctions | Count of roads connected | | Maximum max speed of roads connected |
| 图表, 散点图  描述已自动生成 | | | |
| Level of roads connected | | Presence of crossing | Presence of traffic light |

##### **Street View Features**

This section presents the spatial distribution of the four typical streetview elements, buildings, sky, trees and roads corresponding to different road junctions, to initially illustrate the visual perception characteristics of junctions. In general, sky and road elements see a relatively even distribution on the Camden scale, though for sky elements there are isolated cases with a significantly higher proportion, and various road classes may contribute to the local differences of road elements. The proportion of building elements is commonly higher in all junctions, and shows a decreasing trend from south to north across Camden; conversely, the proportion of tree elements in the streetscape shows an increasing trend from south to north.

Based on the above characteristics, in general, the junctions in the north of Camden may appear more open in visual perception, with a relatively high proportion of natural elements such as sky and trees; the junctions in the south appear more compact, with a relatively higher proportion of artificial elements.

|  |
| --- |
| 图形用户界面, 图表, 散点图  描述已自动生成 |
| Figure 5 |

##### **Activity Features**

The section presents the count of different types of POIs within a radius of 50m around road junctions. Generally, there is a higher frequency of shops, café and fast food POIs appearing around the road junctions, and they are followed by theatres and libraries, the cultural facilities and pubs and bars, the leisure facilities. The schools and hotels are ranked last. Most POIs are concentrated on Holborn and Kingscross and Camden Town, which are the long-term commercial and traffic centres in Camden, and typical attractions for traffic activities.

|  |
| --- |
| 图形用户界面, 图表, 散点图  描述已自动生成 |
| Figure 6 |

##### **Road Network Features**

The calculation results of NAIN and NACH based on the radius of 800m, 1600m and 3200m, for the Camden road network are shown in Figure 7. Different analysis radii enable a multi-scale observation of the attractiveness of the junction to urban activities and the importance of junctions as traffic hubs in the road network.

For NAIN values, NAIN 800 analysis reveals the locations of highly attractive junctions at the local scale. Most of them coincide with the high streets around communities and with densely distributed POI points. The integration of r1600 and r3200 reveals the global difference in the spatial distribution of urban activities at the borough scale. There are significantly more highly integrated road sections and junctions in the south of Camden than in the north, corresponding to the prominent commercial, cultural and service functions of the south of Camden and the central area of London.

For NACH, the analysis results at different scales share more similarities. Junctions with high NACH values ​​are concentrated on most major roads, specifically, most primary roads, A and B roads; and this concentration appears more significant as the analysis radius increases.

|  |
| --- |
| 图表, 地图, 散点图  描述已自动生成 |
| Figure 7 |

##### **ANOVA Analysis**

It is found that, except for the proportion of sky features in the street view and the maximum speed limit of the junctions, there are significant differences in the rest of the built environment features between junctions. Common rules include: high-risk junctions see a greater number and variety of urban activities around and have higher integration and choice values; high-risk junctions are generally on high-level roads, with larger junction sizes and more complete road facilities. In addition, artificial visual elements such as buildings and roads account for a higher proportion in the street view of high-risk junctions; vegetation elements are more evident in low-risk junctions.

|  | **Feature** | **Mean** | | | **ANOVA(One-way)** |
| --- | --- | --- | --- | --- | --- |
|  | **accident\_count\_level** | **0** | **1** | **2** | **p-value** |
| **Comercial and Public Activities** | POI\_culture | 0.018353 | 0.025843 | 0.028892 | 0.0977 |
| POI\_food\_drink | 0.010251 | 0.022472 | 0.046384 | 0.0 |
| POI\_hotel | 0.003720 | 0.005056 | 0.016509 | 0.0002 |
| POI\_pub\_bar | 0.008433 | 0.014045 | 0.024764 | 0.0757 |
| POI\_school | 0.002646 | 0.002247 | 0.007862 | 0.0 |
| POI\_shop\_mall | 0.006792 | 0.015039 | 0.023041 | 0.0001 |
| POI\_SEI | 0.032407 | 0.072714 | 0.155070 | 0.0 |
| **Road Network Configuration** | NACHr3200m | 0.639930 | 0.731494 | 0.821308 | 0.0 |
| NACHr800m\_ | 0.663756 | 0.740601 | 0.802509 | 0.0 |
| NAINr3200m | 0.378617 | 0.478851 | 0.605618 | 0.0 |
| NAINr800m\_ | 0.420421 | 0.518955 | 0.614143 | 0.0 |
| **Street View** | SVI\_Buildings | 0.337894 | 0.343494 | 0.383566 | 0.04 |
| SVI\_Vegetation | 0.270319 | 0.239845 | 0.195577 | 0.0001 |
| SVI\_Roads | 0.540974 | 0.587063 | 0.617437 | 0.0 |
| SVI\_Sidewalk | 0.284856 | 0.293204 | 0.290859 | 0.0 |
| SVI\_Sky | 0.170051 | 0.159078 | 0.148107 | 0.6259 |
| SVI\_Facilities | 0.006548 | 0.010212 | 0.009707 | 0.0462 |
| SVI\_SEI | 0.825232 | 0.829278 | 0.830562 | 0.0104 |
| **Junction Structure and Facility** | min\_junction\_dis | 0.122907 | 0.132619 | 0.126437 | 0.0024 |
| maxspeed\_max | 0.615079 | 0.619775 | 0.692925 | 0.2002 |
| road\_count | 0.217262 | 0.241798 | 0.266509 | 0.0 |
| all minor\_1 | 0.929563 | 0.759551 | 0.372642 | 0.0 |
| minor + major\_2 | 0.059524 | 0.211236 | 0.533019 | 0.0 |
| all major\_3 | 0.010913 | 0.029213 | 0.094340 | 0.0 |
| with\_crossing | 0.107143 | 0.278652 | 0.594340 | 0.0 |
| with\_traffic\_light | 0.028770 | 0.085393 | 0.372642 | 0.0 |