Using Multi-demension Built Environment Features and Machine Learning to Identify Potential Traffic Black Spots

-- Taking Camden in London as An Example

Abstract

Road traffic accidents have long-term been the main safety threat to urban residents, and identifying the concentration areas of traffic accidents, the traffic black spots, are of great significance in improving the traffic safety 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 models present a high overall accuracy of 0.66, and the recall scores for safe and high-risk junctions are 0.8 and 0.71, respectively. 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. Besides, the road configuration features have presented potential in replacing long-term traffic volume data in traffic analysis.

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

# Introduction

Road traffic accidents have long-term been the main safety threat to urban residents. According to the statistics of the World Health Organization (WHO), an estimate of 1.35 million people die in road traffic accidents every year and in high-income countries, the average death rate is 8.3 per 100,000 population(World Health Organization, 2018). Due to the huge hazard of traffic accidents, the prediction and prevention of accident risks are of great significance.

The study outlines an attempt to train random forest classification models to identify the potential traffic black spots based on the junctions’ built environment features, including visual information from street view images, urban activities, configuration of the road network and the structure and facility of junctions.

# Literature Review

## Definition and Identification of Traffic Black Spot

Generally, traffic black spots refer to concentration areas of traffic accidents on the road (Dereli & Erdogan, 2017). However, the specific criteria and identification methods of traffic black spots can vary greatly, according to the differences in urban form and social characteristics. Objects considered traffic black spots can be road intersections, short road sections or mid-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). For example, China defines urban road black spots as: road sections within 500 meters or intersections 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). In addition, 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 datasets the application of clustering methods is relatively limited.

## Built Environment and Traffic Black Spots

Numerous studies have explored and examined the built-environment triggers for traffic accidents. Most studies focus on road crashes dominated by vehicle traffic, and the types of accidents include vehicle-to-vehicle (Asadi et al., 2022), vehicle-to-pedestrian (Ukkusuri et al., 2012), and vehicle-to-bicycle (Osama & Sayed, 2017), and small geographic areas such as census tracts or travel analysis areas are common study units. The corresponding built environment variables may include population and employment density(Fischer et al., Sternfeld & Melnick, 2013), land use and diversity(Yang & Loo, 2016; Osama & Sayed, 2017), street network configuration(Marshall & Garrick, 2011; Choi & Ewing, 2021), road grade and intersection density( Chen & Zhou, 2016), characteristics of transportation and public services(Osama & Sayed, 2017).

Considering that the impact of the same built environment element on collision accidents may be ambiguous or contradictory in some studies, due to regional differences or incomplete selection of variables, some scholars have attempted to summarise the roles of different environmental characteristics in the accident process. Hakkert and Braimaster (2002) divided the occurrence of collision accidents into three parts: exposure, collision risks and injury probability. On this basis, Guerra and Dumbaugh (2020) argued that users' exposure and collision risks as primary mediums are affected by the built environment factors, thereby increasing or reducing the probability of traffic accidents. The variables associated with exposure may include traffic volume and the number of potential traffic participants (population and employment density); Variables associated with risk include vehicle speed and number of intersections. Land use, street configuration and facilities may have mixed or unconfirmed effects.

In the face of smaller-scale research scenarios, the visual information from Street View imagery is believed to be useful for simulating and reflecting features of the built environment in the perception of drivers , cyclists or pedestrians. 快速发展的计算机视觉技术为精确，高效的提取和分析街景图片信息提供了可能，并推动了一系列探索街景信息与街道事故，安全关联性的研究。街景与犯罪，街景与可骑行性，街景与交通黑点等。

# Research Question

# Objectes and Data

## Research Objects

The research takes London Borough of Camden as the research objects.

|  |  |
| --- | --- |
|  |  |
| Figure 1 Junciton Location | Figure 2 Mapping of Road network(with road class) |

## Data Collection

The research is mainly based on the ‘Road Collision Attendants In Camden’ dataset provided by Transport for London(TfL). The dataset contains 4555 road collisions in Camden during 2015-2019, recording 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.

|  |  |
| --- | --- |
| 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 slightly 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. |

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.

# Methods

## Research Framework

## Measurement

### Road Junctions’ Risk as Traffic Black Spots

|  |
| --- |
|  |
| Figure 2 |

### Road Junciton Features

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Group** | **Indicators** | **Data Source** | **Spatial Scale** | **Measure Scale** |
| **Junction Structure and Facility** | Minimum distance to nearest junctions | OS | Local/Global | 0-1(Min-Max) |
| Count of roads connected | OS | Local | 0-1(Min-Max) |
| Maximum max speed of roads connected | OSM | Local | 0-1(Min-Max) |
| Level of roads connected | OS | Local/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 | OSM | Local | 0-True, 1-False |
| **Street View** | Perception of building pixels | SVI | Local | 0-1 |
| Perception of sky pixels | SVI | Local | 0-1 |
| Perception of mix-vegetation pixels | SVI | Local | 0-1 |
| Perception of tree pixels | SVI | Local | 0-1 |
| Perception of road pixels | SVI | Local | 0-1 |
| Perception of road shoulder pixels | SVI | Local | 0-1 |
| Perception of sidewalk pixels | SVI | Local | 0-1 |
| Perception of fence pixels | SVI | Local | 0-1 |
| Perception of wall pixels | SVI | Local | 0-1 |
| Perception of road shoulder pixels | SVI | Local | 0-1 |
| **Comercial and Public Activities** | Number of cultural POI in 50m | OSM | Local/Global | 0-1(Min-Max) |
| Number of food & drink POI in 50m | OSM | Local/Global | 0-1(Min-Max) |
| Number of hotels in 50m | OSM | Local/Global | 0-1(Min-Max) |
| Number of schools in 50m | OSM | Local/Global | 0-1(Min-Max) |
| Number of shops & malls in 50m | OSM | Local/Global | 0-1(Min-Max) |
| Number of pub & bar POI in 50m | OSM | Local/Global | 0-1(Min-Max) |
| **Road Network** | Normalized Angular Choice R800m | OS | Local | 0-1(Min-Max) |
| Normalized Angular Choice R1600m | OS | Global | 0-1(Min-Max) |
| Normalized Angular Choice R3200m | OS | Global | 0-1(Min-Max) |
| Normalized Angular Integration R800m | OS | Local | 0-1(Min-Max) |
| Normalized Angular Integration R1600m | OS | Global | 0-1(Min-Max) |
| Normalized Angular Integration R3200m | OS | Global | 0-1(Min-Max) |

##### Junction Structure and Facility Measurement

Camden borough范围内不同交叉口的结构和功能特征存在差异。对于结构特征，连接在各个交叉口的道路数量决定了交叉口的基本形态；而所连接道路的道路等级组合，如主路和主路，主路和次要路，次要路和次要路，可能影响交叉口的尺寸和交叉口在道路系统中的重要程度。

此外，研究尝试从交叉口的交通设施构成和交叉口所连接road segment的长度和速度限制中，提取交叉口的功能特征。这些变量具体表征为：交叉口是否有交通灯，交叉口是否设置了crossing, 交叉口所连接道路中的最大限速的最大值，和交叉口所连接道路段的最小长度。

本研究中假设，健全的交通设施可以提高交叉口的安全性；而交叉口所连接道路的较高的限速值和两个相邻交叉口间过近的距离，可能使交叉口的交通环境变得复杂，从而提升事故发生的风险。

##### Street View Measurement

从道路网络中提取了xxx个道路端点，

##### Activities Measurement

##### Road Network Measurement

基于空间句法和自然运动理论，城市道路网络的组构特征在一定程度上可以用于反映和描述城市空间中活动的聚集特性。空间句法理论开发了一系列网络度量指标揭示空间组构和城市活动的内在联系。其中，标准化角度整合度（NAIN），作为网络分析中centrality的变体，用于反映不同道路段在特定空间尺度下对城市活动的吸引力；标准化角度选择度，作为betweeness的变体，侧重于反映车流在道路网络中不同路段的潜在分布特征。研究尝试分别以800m，1600m，3600m为分析半径，计算Camden道路网络中每个路段在不同空间尺度下的NAIN以及NACH;在此基础上，为Camden每个交叉口汇总所连接到路段的NAIN和NACH的算数平均值，以此反映道路交叉口在网络中对城市活动的潜在吸引力以及潜在繁忙程度。

## Random Forest Classifier

研究首先利用四类建成环境变量，基于默认参数分别独立拟合随机森林分类器，预测交叉口的风险等级；通过比较不同分类器的表现，了解不同大类的环境变量对交叉口风险的解释性的整体差异。

Based on the built enviornmnent features above, the research try applying the random forest classifiers to predict and identify the risk levels of juncitons for collision accident. 4 groups of built environment features, in group and together, are fitted in the random forest classifiers as independent variables, and the risk levels of juncitons are the dependent variables. The 5 random forest classifier are trained based on the default hyperparameter settings. By comparing model performance, the study first investigate the overall difference of built environment feature groups in interpreting the risk level of junctions.

在此基础上，研究尝试利用全部四类建成环境变量拟合随机森林分类器，利用GridSearchCV方法，基于不同的参数组合，以accuracy为scoring进行模型调试；对每一组参数组合进行，研究采用kfold方法对训练数据集重新划分为训练集和校验集以进行交叉检验。最后，研究通过SHAP特征重要性分析方法，识别对交叉口风险具有潜在更强解释性的具体建成环境变量。

围绕交通黑点的分类，此前的研究中也有基于随机森林分类器的研究实践。

Tanprasert and his colleagues have explored the method to recognize and classify traffic black spots in Thailand based on street view images(Tanprasert et al., 2020).

A distance-aware image segmentation model was developed to extract information of objects surrounding the road and a fully-connected neural networks was trained to identify the black points. The model has shown an accuracy of 69.91% in classifying black and safe spots in Thailand where recall of black points is 75.86%. The combination of street view information and the classification task of traffic black spots is proved to have sufficient potential.

# Results

## Descriptive Results

In this section, preliminary exploratory analysis is conducted to investigate the accident, junction, POI info and street view info datasets.

### Collision Accidents around Road Junctions

##### Count and severity of Accidents around Road Junctions

According to the accident records, for each year during 2015-2019 about 70% of accidents happened near a road junction(Figure 2). The significantly high proportion indicates a potentially high risk of road junctions as traffic black points. The section also presents the distribution of severity class of collision accidents(Figure 3). It is found that for each year most of the accidents belong to the 'slight' type, and the 'serious' type account for about 1/6 of the slight type. Few accidents can be labelled as 'fatal'.

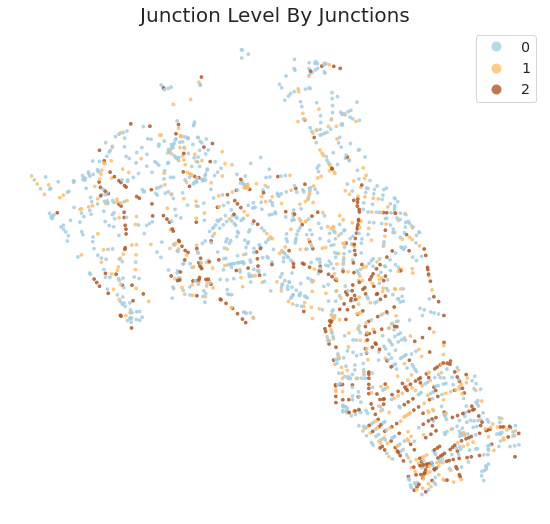
Figure 4 then plots all the road junctions and highlights their difference in accident count. The accident count follows an exponential distribution, where most of the junctions have no accident. Besides, it is found that most of the accidents happened on the junctions near local commercial centres or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

|  |  |
| --- | --- |
|  |  |
| Figure 2 | Figure 3 |
|  | |
| Figure 4 | |

##### Risk Classification of Road Junctions

Based on the count and severity of accidents happening around the junctions between 2015-2019, 1877 road junctions within the Camden borough are divided into three risk levels. 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.

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. There are commonly higher traffic volumes on major roads than on minor roads, which is considered the first cause for the concentration of high risks junctions. However, the distribution of high-risk junctions is usually not even along the same major. More high-risk junctions are concentrated on the road segment in Camden town and Kings cross, Holborn in the south, which indicates that besides the traffic volume, there can be other factors also affecting the accident risks of the junction, such as the urban activities and visual environment characteristics.



### Built Environment Features around Juncitons

This section presents the spatial distribution of different types of built environment features at road junctions within Camden.

##### 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.

|  |  |  |  |
| --- | --- | --- | --- |
|  | | | |
| Minimum distance to nearest junctions | Count of roads connected | | Maximum max speed of roads connected |
|  | | | |
| 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.

|  |
| --- |
|  |
| 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.

|  |
| --- |
|  |
| 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 intersections 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.

|  |
| --- |
|  |
| Figure 7 |

## 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 with all feature groups has better performance than models trained based on a single feature group. Besides, some feature groups have shown a significantly stronger fit than other feature groups in classifying the junctions risk levels. Specifically, models fitted with the activity and streetview characteristics separately have a weak ability in classifying the risk level of accidents. Their accuracy scores on the test dataset are lower than 0.55. In contrast, models fitted with multi-scale space syntax measures and with the junction structure and function have better performance. The accuracy scores of the two models on the test data set are 0.638 and 0.59, respectively. Beyond that, the model trained based on all the feature groups has gained the highest accuracy score of 0.681on the test datasets.

Activity 模型存在拟合不充分问题，其余组别则可能过拟合。

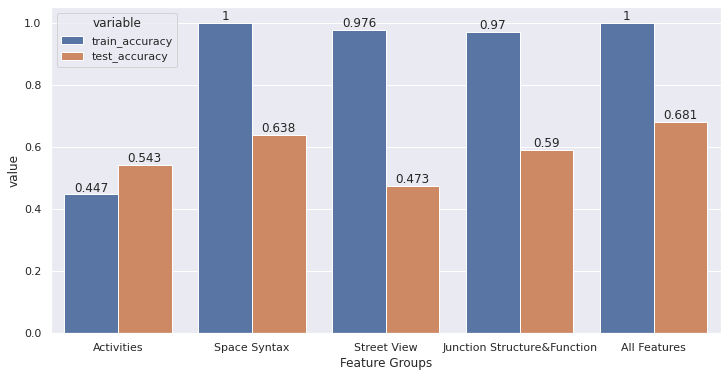


Figure 7

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

|  |  |  |
| --- | --- | --- |
|  | 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 | 350 |

According to the analysis, the best hyperparameters couple are 60 for ‘max\_depth’ and 350 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 0.665 and presents strong abilities in identifying safe and high-risk junctions. Specifically, the model shows the best performance in classifying safe junctions. The precision score is 0.77 and the recall score is 0.83, presenting that model is good at distinguishing both safe and unsafe junctions.

However, in identifying junctions with detailed risk levels, the performance of the model is slightly weaker. Model gains relatively high precision and recall scores in classifying high-risk junctions, but poor scores in classifying low-risk junctions. As the recall score in classifying high-risk junctions is significantly lower than that in classifying safe junctions, it is inferred that the model is inclined to classify a certain number of high-risk junctions as low-risk junctions.

Table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **precision** | **recall** | **f1-score** | **support** |
| **0** | 0.77 | 0.83 | 0.80 | 100 |
| **1** | 0.24 | 0.21 | 0.22 | 39 |
| **2** | 0.72 | 0.69 | 0.71 | 49 |
| **accuracy** |  | | 0.66 | 188 |
| **macro avg** | 0.58 | 0.58 | 0.58 | 188 |
| **weighted avg** | 0.65 | 0.66 | 0.66 | 188 |

### Feature Importance

The figure plots the global importance of each feature in classifying different junction risk levels in the random forest model.

The global importance of multi-scale configuration features, and the structure and facility features of junctions, are generally higher than that of street view and activity features in the model. Specifically, whether the junction is connected only by minor roads ranks first as the most important feature; the features ranking from 2nd to 6th are all road configuration features, namely NAINr3200，NAINr1600， NACHr3200, NAINr800 and NACHr1600, where integration measurements can weigh more heavily than the choice measurements; features after NACHr1600 have seen a significantly lower global importance in the model, and the activity features can have poorer performance than that of street view features.

In addition, the top 6 features in the global importance ranking show better performance in predicting safe and high-risk junctions. Featuring maintaining comparative advantages in predicting low-risk junctions are concentrated in the middle and lower sections of the importance ranking, such as the proportion of the road and road shoulder elements in the streetview, the minimum distance from junction to the adjacent junctions. These features are of low global importance in the model, which may explain why the model overall has weak performance in classifying low-risk junctions.

图表

描述已自动生成

# Discussion and Conclusion

**在模型预测层面，**

* + 基于空间句法变量单独拟合的分类模型的预测较为充分；
  + 单独采用视觉，活动或交叉口的结构和功能信息对风险的预测能力弱。
  + 对比表现最好的单一模型(基于空间句法measurements的模型)，综合模型的表现有小但明显的提升，表明其他环境信息和要素同样对提升模型进度有所贡献。尽管其他建成环境信息可能与空间句法变量包含的信息存在重合。
  + 经历参数调整后，综合模型在检验数据集上的精确度是 0.66，对安全和高危的交叉口有良好的分类表现，但是对低风险交叉口的分类能力较弱。
  + 对低风险交叉口分类能力弱的原因可能是对于低风险交叉口的分类规则较为模糊且本研究中涵盖的低风险样本总数较少，模型对对应特征的学习不充分。
  + 考虑到，构建预测模型的重要目标是在当下的城市管理和未来的城市道路设计中识别潜在的**高风险交叉口** ，减少和预防伤亡事故，尽管存在缺陷，模型整体上是有效的。

**模型解释层面**

* 交叉口是否仅基于minor road，以及交叉口的多尺度整合度和选择度特征，都在特征重要性排序中占有首要位置。其次的重要性排序整体为交叉口结构和设施特征，交叉口的视觉信息和交叉口的活动。
* 在本模型中，表征和包含大尺度或系统性空间联系的建成环境要素，如3200m为分析半径的整合度和选择度，参与构成交叉口的道路等级等，重要性整体强于仅表征局部特征的建成环境要素，如交叉口不同POI的数量，树木，墙体等视觉要素 。
* 特征重要性指向特定模型，只表征构建模型时哪些特征更有用。它们不应被解释为预测变量和目标之间的直接依赖关系。
* Shap values不同于Permutation Based Feature Importance特点的讨论

**交通风险与建成环境**

* 道路交叉口的交通风险或更应被视为一种**系统性**问题加以分析，预测和治理。可以从两个方面认知
* 其一，从本研究得到的证据看，基于单一类别建成环境变量拟合的模型中，基于多尺度道路组构特征的模型表现最好；基于全部变量的模型中，表征全局特征的变量重要性整体优于表征局部特征的。这似乎表明，变量中所捕获和表征的涉及道路和城市系统的全局信息越多，对交叉口风险的预测能力越强。
* 其二，基于Guerra and Dumbaugh (2020)的理论，交通事故的发生与建成环境中的暴露和风险两类因素有关。其中暴露相关要素，例如潜在的交通流量和人口密度，应当被视为一种系统的全局或跨尺度特征；而风险，例如限速和交叉路口，则应被视为一种局部特征。如果将基于空间句法方法的道路网络分析变量（e.g.NACH）视作长期交通流量分布或Vehicle miles traveled (VMT)（而非瞬时流量）的有限替代，可以看到代表暴露的相关特征整体在交叉口事故风险中发挥了更大作用。
* 如果我们的研究对象并非只有交叉口，那么与风险关联的要素可能会更重要，或者产生更多元的结果；在只包含交叉口的情况下，交叉口特征可能更多服务于道路系统的整体特征，一些局部建成环境的影响减弱。

**其他贡献：**

* 研究首次将空间句法方法与机器学习结合进行道路交叉口的风险预测和分类。并在特征重要性分析中检出了基于空间句法的多尺度道路组构特征的重要作用。一方面，这开拓了空间句法方法的应用场景；另一方面，这为当前道路交叉口的管理和维护，以及未来道路网络和交叉口的设计提供了借鉴。
* 在本研究中，不同于此前Tanprasert et al（2020）的研究，由纯粹街景信息拟合的分类器在交叉口风险分类中的表现不佳。可能是由于研究： 1.仅以交叉口为研究对象，街景特征存在趋同性。2.对于交通黑点的定义和分类方法存在差异。尽管如此，在综合模型中，街景信息中road, roadshoulder, building, sky 等对于分类交叉口风险仍具有较为重要的贡献。本研究初步检验了此前研究方法和结论的可靠性，相关方法在不同场景仍有发展前景。

**局限：**

* 尽管本研究设置了3个交叉口风险层级，但能够实现有效分类的只有安全交叉口和高风险交叉口。这主要是由于目前的事故数据在Camden Borough范围内较为稀疏。 模型对低风险交叉口的特征的学习不充分。

Table 1. Table caption

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |

# Acknowledgements

# References

In the draft, if you want to cite a paper, please put the paper in our shared folder in Mendeley, and then cite the paper from this folder.

If you use Mendeley, you can insert the references using the following commands in word: References -> Mendeley -> Insert Bibliography.

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.