Development of a street-by-street analysis of spatial factors in street crime using publicly available crime data by census block

Atsushi Takizawa* and Yoko Tanaka

Graduate School of Human Life and Ecology, Osaka Metropolitan University, Japan

*Corresponding author: takizawa@omu.ac.jp

Abstract

This study investigates street crime patterns in Osaka City, Japan, by analyzing spatial factors and identifying specific spatial attributes associated with crime occurrences at a resolution finer than census tracts. Although Japan is renowned for its low violent crime rates, evolving social conditions underscore the importance of continuous crime research. Publicly available crime data in Japan, typically aggregated by census units rather than individual crime points, pose challenges for granular spatial analysis. To address this, the study introduces a method that focuses on street segments By generating road topologies from street polygons and incorporating 3D building data alongside pedestrian flow derived from mobile data. a graph convolutional network (GCN) was developed to predict crime occurrences across street segments. Using data on theft of a bicycle and theft from a car, the GCN achieved accuracy comparable to traditional machine learning methods based on census tracts, with a goodness-of-fit exceeding 0.8 for estimated crime distributions. The findings highlight the critical role of micro-spatial features, such as street width and pedestrian flow, in shaping crime dynamics and suggest practical implications for crime prevention through environmental design (CPTED) strategies.

Key words: street crime, street segment, graph convolutional network, integrated gradients, human flow data, Plateau

1. Introduction

This study aims to develop a model that explains street crime patterns in Osaka City, a major urban center in Japan, by utilizing spatial factors and identifying the spatial attributes associated with crime occurrences. This model employs a graph-based approach with a resolution at the street segment level, which offers finer granularity than the traditional census tractunit commonly used for crime data aggregation. The following sections provide the background and rationale for this study.

Japan is frequently regarded as a safe country with notably low levels of violent crime compared to global standards. For instance, according to United Nations statistics for 2020 (United Nations Office on Drugs and Crime, 2020), Japan ranked fifth lowest in murders per capita among 134 countries with recorded statistics. However, sociological factors such as the rapid increase in single-person households (Osborn et al., 1992) and growing racial and ethnic diversity (Sampson & Groves, 1989) are known to influence crime occurrences. These shifts underscore the need for continuous research, as the persistence of Japan's current safety levels cannot be assumed indefinitely.

It is widely acknowledged that crime tends to cluster in specific locations rather than occurring randomly in urban environments, a foundational concept in environmental criminology (Andresen, 2014). the spatial scale of analysis plays a critical role in understanding these patterns. For instance, Cohen (1980) demonstrated that while street prostitution in New York City showed no significant variation at the regional or census tract levels, notable differences emerged when examining the micro-scale of city blocks and street corners. Such evaluations are crucial when considering crime prevention strategies like crime prevention through environmental design (CPTED). Originating from the work of Jacobs (1961) and Newman (1972), CPTED has been widely adopted across countries and is built on principles such as territoriality, surveillance, access control, activity support, and target hardening.

Among these, natural surveillance is a central concept, asserting that the presence of people observing public spaces can deter criminal activity.

However, the relationship between environmental features and crime is complex and sometimes contradictory. For example, some studies suggest that areas with high population density or well-maintained streets may be more vulnerable to crime (He & Li, 2021; Lee & Contreras, 2021). Conversely, while certain research argues that green spaces increase crime by obstructing visibility (Kim & Hipp, 2018; Maruthaveeran & Bosh, 2015), others posit that green spaces reduce crime (Wolfe & Mennis, 2012). These inconsistencies highlight the challenge of creating a universal model for explaining crime, as crime dynamics are often context-dependent. As noted by De Nadai et al. (2020), it is essential to tailor crime analyses to the specific characteristics of each region to account for its unique socio-spatial dynamics.

To examine the relationship between crime occurrences and spatial characteristics, it is essential to access crime data with sufficient spatial resolution. However, the United Kingdom is the only country that provides detailed crime occurrence point data as open data at a national level (DATA.POLICE.UK., 2013). Other examples of cities offering detailed crime data include a limited number in Europe and North America, such as NYC Open Data (2013), Chicago Data Portal (2012), and Toronto Police Service (2022). In most other countries, crime data is typically aggregated and published by government entities or organized by mesh units. In Japan, the National Police Agency (2018) publishes nationwide open data on street crime, but the spatial unit is limited to national census units. Conducting detailed spatial analyses in areas where point data is unavailable requires experts to request granular data directly from police or other organizations. However, in Japan, there is no standardized procedure for obtaining such data, and requests are often denied. As a result, microscopic research on crime at the street scale or finer is limited in Japan (e.g., Ishikawa et al., 2009; Takizawa, 2013; Kishita et al., 2016; Hashimoto et al., 2023).

Against this background, this study proposes a method for analyzing the relationship between street crime and spatial characteristics by utilizing open crime data aggregated at the census tract level but analyzing it at a finer granularity—the level of street segments. Specifically, road topology is generated from street polygons using image processing techniques and then divided into polygons for individual streets. detailed 3D building data is incorporated to estimate the functional volume of each street based on its use. Furthermore, pedestrian and vehicular flow data, derived from smartphone-based point-type people flow information, is analyzed to capture transport-specific movement patterns for each street segment. A graph convolutional network (GCN) is then employed, using these street level characteristics as explanatory variables, to estimate the number of crimes occurring on each street segment. the data is subsequently aggregated by small areas, and the model is trained to ensure that the total number of crimes aligns with the aggregated crime data. the generalization performance of the GCN is compared to that of a machine learning model based on general small areas. Finally, the number of crimes is visualized at the street level to investigate the underlying causes of crime.

In the following chapters, the structure of this study is presented: Chapter 2 reviews prior research to clarify the positioning and novelty of this work. Chapter 3 explains the proposed method, and chapter 4 introduces the data used. Chapter 5 validates the proposed method, and chapter 6 discusses the results. Finally, chapter 7 concludes with a summary of the study's findings.

2. Literature review

This research, like many others, builds on the theoretical foundations of environmental criminology. Environmental criminology originated with the routine activity theory (Cohen & Marcus, 1979) and the rational choice theory (Cornish & Clarke, 1986), eventually evolving into the crime pattern theory (Brantingham & Brantingham, 1993). Crime pattern theory posits that crime emerges from the synergistic interactions of a criminal, the target of the crime, and specific temporal, spatial, and environmental conditions.

With the advancement of geographic information systems (GIS) and the increasing availability of spatial data, numerous empirical studies have been conducted based on these theoretical principles. to position this research within the existing body of

work, this study organizes the literature around three key points: (1) the types of data utilized, (2) the spatial structures employed to model crime patterns, and (3) the analytical methods applied in these studies.

Over time, the types of data used in crime analysis have expanded significantly, encompassing socio-economic environments, land use, street network structures, and population distributions. Socio-economic data, which describe macro-community spatial attributes such as population distribution, poverty levels, and racial composition, have been central to crime research for decades (Shaw & McKay, 1942; Curry & Spergel, 1988; Sampson et al., 1997; Graif & Sampson, 2009). Similarly, Land use studies have investigated crime patterns for a long time, with seminal work indicating that crime is more likely to occur in areas where people congregate, such as restaurants or commercial districts (Roncek & Maier, 1991). Recent studies have gone further, exploring the relationship between crime and building use, green spaces, and other spatial attributes (Sypion-Dutkowska & Leitner, 2017; Saraiva & Teixeira, 2023).

Research on CPTED has focused on detailed spatial elements like streetlights, surveillance cameras, and the visibility of public spaces, all of which have demonstrated significant relationships with crime occurrence (Welsh & Farrington, 2008; Deisman, 2003; Takizawa, 2013; Lee et al., 2019). However, the availability of such data is often limited. In response, studies leveraging artificial intelligence (AI) have begun to classify spatial components from street view images using semantic segmentation techniques (Hipp et al., 2021; Xie, 2022). Street topology has also become a major focus in crime analysis. Crime pattern theory links criminal behavior to the network structure of streets, modeled as nodes and edges. Graph-based features like integration values and betweenness centrality, originating from space syntax (Hillier & Hanson, 1986), are frequently used to describe street networks (Summers & Johnson, 2017; Marchment et al., 2021; Xian et al., 2023). Additionally, human flow data derived from mobile devices provide dynamic measures of ambient populations, offering more nuanced insights than static census data (Andersen, 2011; He et al., 2020).

Spatial statistical models, which account for spatial autocorrelation (Anselin et al., 2000), remain widely used due to the geographical nature of crime distribution. However, with advances in AI and machine learning, models that capture nonlinear relationships while improving explanatory power and prediction accuracy have gained prominence. gradient boosting methods are now mainstream in machine learning and are increasingly paired with SHAP (Lundberg & Lee, 2017), an interpretability tool, in crime research (Deng et al., 2023; Kim & Lee, 2023). GCNs, a deep learning technique, have recently been applied to crime prediction models. Unlike traditional methods, GCNs integrate the feature values of street segments and surrounding network topology directly into the model (Zhang & Cheng, 2023; Gu et al., 2024).

This study contributes to the literature in several novel ways. First, it uses crime data aggregated at the census tract level but conducts analysis at the street level, achieving higher spatial resolution than previous studies. Second, it leverages comprehensive land use data to understand the characteristics of streets. While traditional studies often rely on points of interest (POIs) for this purpose, their lack of comprehensiveness limits their utility. Instead, this research utilizes the detailed 3D city model "Plateau" (Ministry of Land, Infrastructure, Transport and Tourism, Japan, 2020) to comprehensively digitize building use along streets. Unlike prior research that relies on aggregated population data, this study uses point-based pedestrian flow data estimated by travel mode obtained from smartphones. This approach enables a more granular understanding of human movement dynamics along street segments. Additionally, while GCNs have been used for crime prediction models, their application to descriptive models is rare due to the interpretability challenges of deep learning. To address this, the study employs Integrated Gradients (Sundararajan et al., 2017) and SHAP to provide transparent explanations for model outputs, making this a significant methodological advancement.

3. A model that explains the occurrence of crime in each street (street segment model)

The idea of the street segment model is to estimate the number of crimes in each street in a small area, take the sum of these numbers, and approach the number of crimes in the open data. For this purpose, the model is a graph convolution network model

with the edges of the street network as spatial units. The model is described below.

3.1. Construction of neighborhood graphs

Since the number of crimes for individual street units are unknown, the model cannot be trained using a standard supervised learning framework. To address this, the number of crimes is initially estimated for each street, and the weights of the GCN are optimized so that the sum of these street level estimates closely matches the actual number of crimes recorded in the corresponding small area.

The first step in this process is to associate each small area with its corresponding streets. Let A be the set of small area polygons, E be the set of street segment polygons, and $E_a(\subset E)$ be the set of street segment polygons belonging to small area $a \in A$. As street segment $i \in E_a$ in $a \in A$ may extend into multiple small areas, such street polygon is split with the relevant small area polygons to obtain the ratio to its area, which is denoted by $r_{i,a} \in (0.0,1.0]$. For training the GCN model, the graph is constructed in units that include each street segment and its surrounding neighborhood, referred to as the neighborhood graph. As shown in Figure 1, the neighborhood graph for a central street segment $i \in E$ belonging to small area $a \in A$ and the edges are expanded to $d \in \{0,1,2,\dots\}$ by width-first search is defined as the neighborhood graph of i, and its edge set is defined as $E_{i,d}(\subset E)$. d is given as a constant in advance.

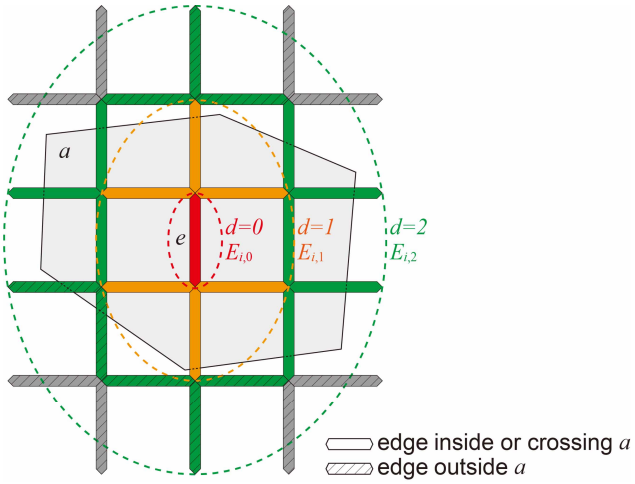

Fig. 1. A neighbor graph.

3.2. GCN model

Figure 2 illustrates the proposed GCN. The implementation was carried out using PyTorch (2016) and PyTorch Geometric **(**PyG**)** (2018). When the graph data for the neighborhood graphs of each edge in a small area is input into the GCN as a batch, the model estimates the number of crimes occurring on those edges. Since standard graph convolutional methods typically assume node-based data with attributes assigned to nodes, the input data is preprocessed into a dual graph where the edges are treated as nodes. However, for the sake of simplicity, the following explanation assumes that the edges retain their original representation.

The input data is a matrix ($\boldsymbol{X}$) where the rows represent explanatory variable vectors for the central and surrounding edges of the neighborhood graphs in a given small area. These vectors are stacked in the row direction for all neighborhood graphs. Additionally, data related to the graph topology is included. Before the learning process begins, The values in $\boldsymbol{X}$ are standardized across the entire dataset to ensure numerical stability during optimization. Various methods for graph convolution were considered. After comparing typical approaches, this study adopted Residual Gated Graph ConvNets (Xavier and Thomas, 2017) for their performance and suitability.

Let C be the set of modus operandi of crime, $N_i \subset E$ denote the set of edges adjacent to edge $i \in E_a$ and contained in $E_{i,d}$, anrzd let $x_i \subset \boldsymbol{X}$ represent the attribute vector of i. This graph convolution updates x_i to x_i' as follows:

$$x_i' = \boldsymbol{W_1} x_i + \sum_{j \in N_i} \eta_{i,j} \odot \boldsymbol{W_2} x_j \tag{1}$$

$$\eta_{i,j} = \sigma\left(\boldsymbol{W_3} x_i + \boldsymbol{W_4} x_j\right) \tag{2}$$

In the above equation, $\boldsymbol{W_1} \sim \boldsymbol{W_4}$ is the weight matrix, $\odot$ is the vector's element-wise multiplication operator, and σ is the sigmoidal function. The convolved feature matrix is normalized by group normalization and passed through the rectified linear function (ReLU). This is made into one module and used as the model with $d = 1$, and this module is increased by one for each additional neighborhood. Since only the feature values for the central edge of each graph are required, the vector is extracted by IndexSelect, converted to a scalar by Liner, and passed through Softplus, which returns non-negative real numbers, and finally, the number of crimes at the edge i as $\hat{n}_{c,i}$ is estimated where $c \in C$.

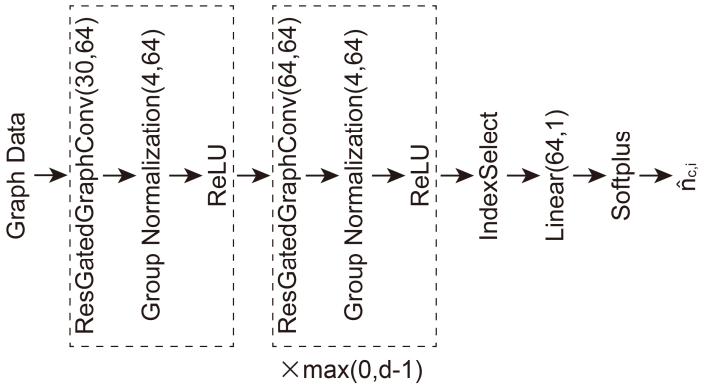

Fig. 2. Structure of the GCN.

The loss function employed in this study is based on the squared error between the actual number of crimes $N_{c,a}$ in small area $a \in A$ for crime $c \in C$ and the estimated value, weighted by the proportional contribution of each street segment. Although a Poisson loss function would typically be more appropriate for frequency-based data sets like this one, the learning process did not converge effectively when using Poisson loss. As a result, the squared error loss function was adopted instead.

$$loss_{c,a} = \left(N_{c,a} - \sum_{i \in E_a} r_{i,a} \cdot \hat{n}_{c,i} \right)^2 \tag{3}$$

4. A model that explains the occurrence of crime in each small area (small area model)

A regression model was constructed to predict the number of crimes in each small area and evaluate the performance of the proposed GCN model. This approach reflects the standard methodology typically applied when analyzing this type of crime data. The explanatory variables for this regression model were derived from the edge attributes described in Chapter 5.4 by aligning them with the small areal polygons. For street segment data, the lengths were aggregated using sums, and the widths were summarized using means. For the estimation model, a general machine -learning framework for regression problems was employed. PyCaret (Ali, 2020), a platform that facilitates simultaneous comparison of multiple state-of-the-art machine learning models, was used for validation. Among the tested models, the gradient boosting regressor (GBR**)** (Friedman, 2001), an ensemble learning method based on decision trees, was selected for its performance and suitability for the task.

5. Data

This chapter describes the data used and the processing methods used.

5.1. Target area

Figure 3 illustrates the map of Osaka City, the target area for this study. Osaka City comprises 24 wards, which can be broadly categorized into two distinct regions based on their geographical and functional characteristics. The central area includes six wards with a high daytime population, characterized by major railroad terminals, offices, and commercial facilities. Surrounding the central area are the peripheral wards, encompassing 18 wards where residential neighborhoods and factories coexist. Given the significant differences in activity patterns between these two regions, the analysis was conducted separately for The central and surrounding areas. The central area covers an area of 38.0 km² and has a nighttime population of 585,944, while the surrounding area spans 182.3 km² with a nighttime population of 2,166,468 (Statistics Bureau of Japan, 2021).

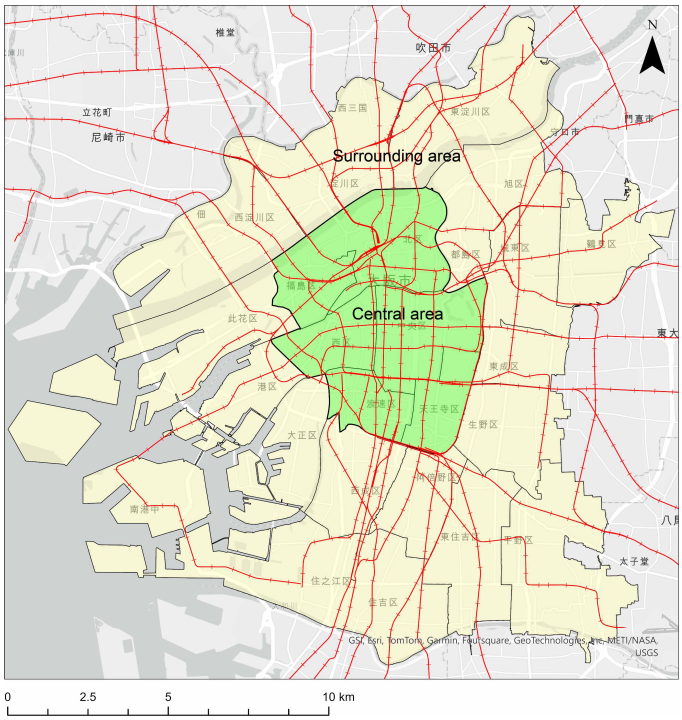

Fig. 3. Central and surrounding areas of Osaka City. Red lines indicate railroads.

5.2. Crime data

The number of crimes by modus operandi over the five-year period from 2018 (the first year data became available) to 2022 was obtained from the Crime Open Data Site (Osaka Prefecture Police, 2023; see Table 1). During this period, there was a general downward trend in crime until 2021, partly attributed to the COVID-19 pandemic. However, in 2022, the number of crimes increased by 12.1% compared to the previous year. Among the reported crimes, bicycle theft (**TB**) was the most prevalent, followed by theft from a car **(TC)**. For this analysis, data from 2018 and 2019 were combined because the human flow data (described later) pertains to 2019, a period before the Covid-19 pandemic. Only TB and TC crimes, which have sufficient incident counts for meaningful analysis, were included in the study. The number of TB and TC incidents by small area was visualized using the basic unit district polygons of the 2020 census (see Figure 4). The map includes 1,904 small area polygons, with 541 in the central area and 1,363 in the surrounding area. The mean area of these polygons is 0.116 km², with a standard deviation of 0.123 km². The visualization depicts the sum of incidents for 2018 and 2019. Geographically, TB incidents were more concentrated in the northern and southern

parts of the city center and in surrounding areas near railway stations. In contrast, TC incidents were primarily concentrated in the southern part of the city center.

Tables A3 and A4 in the supplemental material presents the total number of crimes by location for TB and TC in both the central and surrounding areas during 2018 and 2019, respectively. The data indicate that parking lots and streets are the most common locations for both types of crimes. While The total number of crimes in the surrounding area is approximately double that of the central area, the crime density (crimes per unit area) is significantly higher in the central area.

Table 1

Number of incidents by type of operation in each year for the entire city of Osaka.

Modus operandi	2018	2019	2020	2021	2022
Theft of a bicycle (TB)	**12,512**	**12,123**	9,430	9,000	10,541
Theft from a car (TC)	**3,704**	**2,511**	2,132	1,743	1,791
Theft of car parts	1,452	1,259	1,029	1,188	779
Theft from a vending machine	1,028	560	138	130	211
Theft of motorcycles	689	544	507	427	335
Theft of automobiles	484	273	311	218	190
Bag snatching	229	154	106	64	106
Total	20,098	17,424	13,653	12,770	13,953

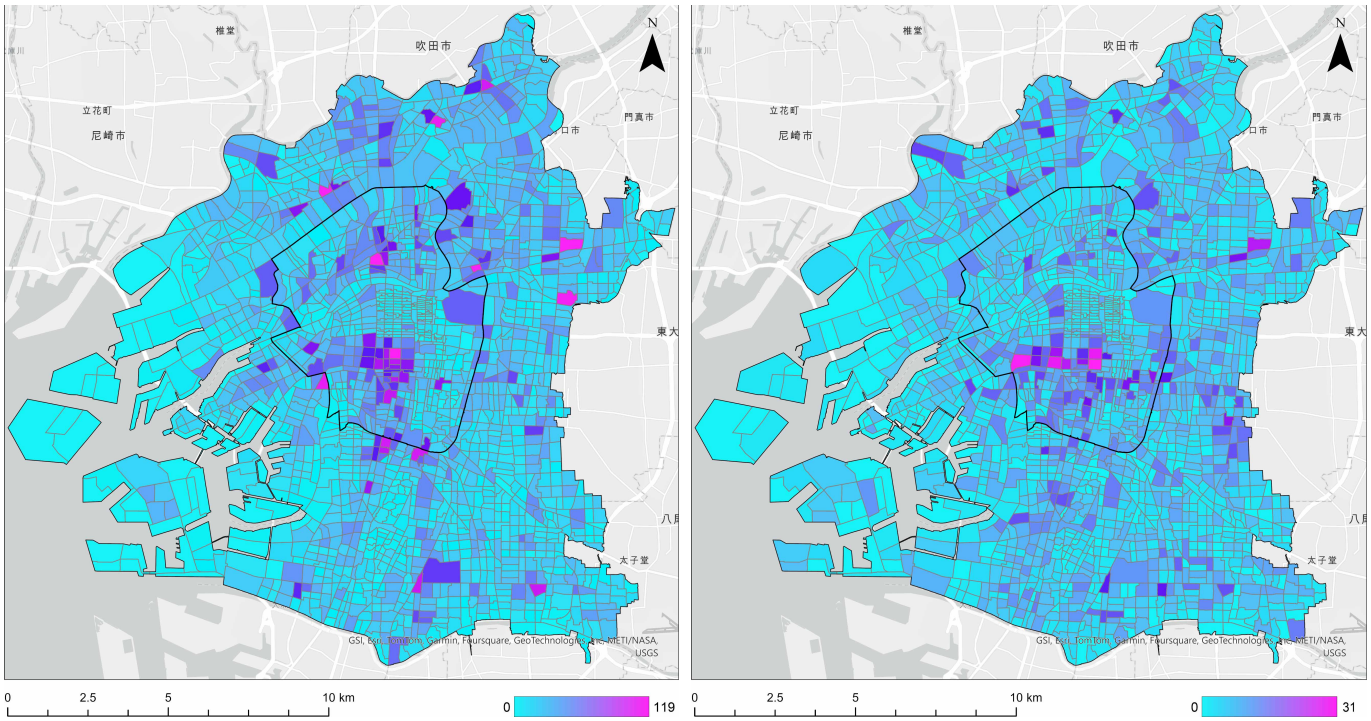

Fig. 4. Number of TB (left) and TC (right) by small area from 2018 to 2019.

5.3. Street data

The street data used in this study requires two types of attributes: polygonal attributes for expressing the general shape of the street, such as its width, and network attributes for inputting into GCN. The process of constructing a large-scale street network covering the entire Osaka City area was automated wherever possible. Street polygons were generated by merging road segments while excluding expressways, toll roads, and other non-relevant features. This data was derived from the ESRI Detailed Map 2020 (Kinki region). The area of these street polygons was rasterized as a binary image, and then the street network was created from the

raster data after performing a thinning process. Furthermore, the area each street edge occupies in the segment polygon was determined by calculating the Voronoi diagram with obstacles for the street boundary (see Figure 5). As a result, 105,489 street segments were obtained for the entire city of Osaka, and of these, 15,228 streets in the central area and 90,261 streets in the surrounding area were used as central edges. Due to the inherent properties of Voronoi diagrams, the areas around intersections are generally not rectangular. However, for the purposes of this analysis, the length and width of each street are required. the method described in Appendix B was employed to approximate these dimensions.

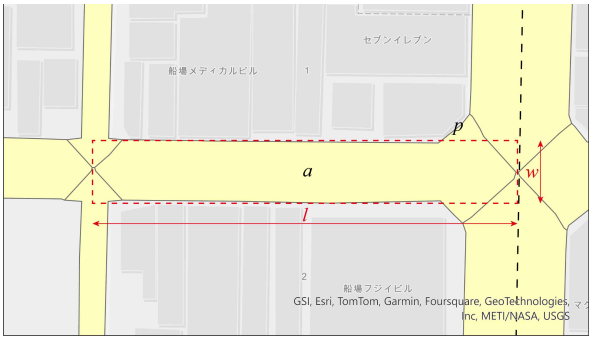

Fig. 5. Image of street polygons.

5.4. Explanatory variables

We characterize each street using explanatory variables that may be related to the occurrence of street crime.

5.4.1. Area by building use

Building use is a key explanatory variable in urban analysis, as supported by previous studies. This study utilized Plateau 3D building data for Osaka City (2020), which provides detailed three-dimensional information on building uses. Considering that convenience stores and large buildings are often set back from streets, buildings located within a 30-meter buffer zone from street polygons were associated with the streets in this analysis. Using this approach, 600,675 buildings were linked to streets out of the 615,938 total buildings in Osaka City. Next, the total floor area of the nearest buildings for each segment polygon was calculated based on their detailed use. While the Plateau 3D dataset classifies buildings into 60 distinct use types, these categories were too granular for this analysis. To address this, the detailed uses were aggregated into 21 broader use categories, as shown in Figure 6. The total floor area for each of these 21 categories was then tabulated and used as an explanatory variable for the analysis.

Detached house, row house, apartment house, apartment house with store, residence with store, residence with factory, commercial and sales facility, restaurant, business facility, service entertainment and recreation facility, educational facility, cultural facility, religious facility, medical and welfare facility, lodging facility, industrial facility, supply and processing facility, transportation and communication facility, government facility, other facility, unknown.

Fig. 6. Building use classification.

5.4.2. Open space-related area

Using the Plateau 3D data, three primary types of open spaces were targeted for analysis: parks, flat parking lots, and vacant lots, which represent the main categories of open spaces in urban environments. These Open spaces were analyzed by calculating their total area while allowing for overlaps among open space types. To establish their spatial relationship with streets, only open spaces located within 1 meter of each street polygon were included in the analysis.

5.4.3. Human flow data

The analysis utilized point-based flow population data provided by Agoop Inc**. (**2019**)**, which consists of location information collected from smartphone applications. Only data points located on the street polygons were extracted for analysis. To identify travel modes, we modified the methodology from the previous study by Takagi and Takizawa (2019). The data were categorized into three travel modes: pedestrians (8,571,405 points), bicycles (1,741,332 points), and automobiles (1,962,386 points). Since five of the ten days of data for each season were holidays, the number of points for holidays was adjusted to 2/5 of the original number to ensure consistency in the analysis. Figure A1 in the supplemental material illustrates the distribution of pedestrians, bicycles, and automobiles across various parts of the city, highlighting higher densities in the city center (especially in the northern and southern regions) and the surrounding areas. For the analysis, the human flow data were divided into daytime (6:00–18:00) and nighttime (0:00–6:00, 18:00–24:00) periods to account for temporal variations in activity patterns.

5.4.4. Others

The XY coordinates of the centroid of the spatial unit were used. In addition, the length and width of each street obtained in 5.3 were used.

5.5. Basic statistics and correlation of variables

A total of 34 explanatory variables were included in the analysis. The basic statistics for these variables are presented in Table A1 in Appendix A of the supplemental material. Additionally, to examine the correlations between variables, including the number of crimes, correlation coefficients were calculated for each small area and are presented in Table A2. To prepare the data for analysis, the street polygons were assigned to the small area with the largest overlap. For each small area. The street polygons are assigned to the small area with the largest overlap, the street length is the sum of the street lengths, and the street width is the average width. The center of gravity of the street polygons is changed to the center of gravity of the small area polygons. The small area data compiled in this way is used in the small area unit model.

6. Verification

The model is divided into the central and surrounding areas shown in Figure 3, and validation is performed for both the small area and street unit models.

6.1. Learning and evaluation settings

After evaluating several optimization methods for the GCN model, RAdam (Liu, 2020) was selected as the optimizer. The following hyperparameters were applied: $epoch = 250,\ batch\ size = 1,\ decoupled\ weight\ decay = \text{True},\ weight\ decay = 1.0e-4$. The learning rate is shown separately in Table SXX because the accuracy of the model varied depending on the type of crime and the complexity of the GCN model. Next, an accuracy evaluation was conducted with the hyperparameters fixed in this manner. To evaluate the accuracy, we estimated the number of TBs and TCs using data from 10-fold cross-validations in which the training and validation data were divided so that the training and validation data were the same combination of sub-areas. First, we compare the accuracy (R2, RMSE) of the street unit model by varying $d \in \{0,1,2,3\}$. In the deep learning model, it is necessary to set the appropriate number of epochs by Early stopping. Therefore, the number of epochs for a particular fold was set to the epoch where the average value of R2 for each epoch of the validation data for the other nine folds was the largest value. In obtaining the curve of the mean values of the nine folds, we used local regression to smooth out minor precision deviations from epoch to epoch.

Finally, for each crime/area data, the model with the highest accuracy, d, we again find the epoch with the largest R2 by local regression with 10 folds, train the final model with full data up to this epoch, and analyze the results in detail.

The GBR model was parameter-tuned beforehand by using the PyCaret function with 10-fold cross-validation, and its accuracy was evaluated based on the test data. Finally, the model is optimized on the full data set, and a final model is created to explain it.

6.2. Accuracy results

Table 2 presents a comparison between the GCN model with the highest accuracy and the GBR using cross-validation. For the GCN models, the value of d (depth of the neighborhood graph) corresponds to the configuration that achieves the highest accuracy for each specific model. Comprehensive results for all models are provided in Table A5 in supplemental materials. For bicycle theft (TB), the GCN model with $d = 1$ achieved the highest accuracy, indicating that crimes in this category are influenced by nearby streets in addition to the street under analysis. In contrast, for theft from a car (TC), no consistent trend regarding d was observed. The GBR demonstrated greater accuracy for TB compared to the GCN model, particularly in urban areas. The higher accuracy for TB may be attributed to the fact that the number of TB incidents is approximately four times greater than that of TC. However, the average accuracy across cross-validations did not achieve statistical significance. Despite this, the GCN model provides estimates on an edge-by-edge basis, which introduces challenges in model adjustment and accuracy optimization. This ability to analyze street-specific influences is a notable result, highlighting the model's potential utility despite its complexity.

Table 2

Average accuracy of 10-fold cross-validations of two models.

Modus operandi	Area	GCN			GBR	
		d	R2	RMSE	R2	RMSE
TB	Center	1	**0.700±0.130**	**9.183±1.963**	0.677±0.157	9.512±2.165
	Surroundings	1	**0.550±0.145**	**8.891±1.921**	0.537±0.133	8.972±1.701
TC	Center	2	**0.522±0.171**	**2.887±0.384**	0.496±0.127	2.989±0.357
	Surroundings	0	**0.317±0.097**	**2.328±0.143**	0.301±0.074	2.360±0.111

6.3. Evaluation of crime locations

The next step involves verifying the certainty of crime locations for each street. The crime data utilized in this study includes information about the locations where each crime occurred, as detailed in Table 3, as well as additional location information. Since the locations of crimes occurring at residences correspond to the spatial data employed in this analysis, these crime occurrences were limited to those locations for confirmation purposes. Let T represent a set of three crime occurrence locations listed in Table 3, and let $E_{a,t} \subseteq E_a$ be a set of streets in small area $a \in A$ that contains $t \in T$. Furthermore, let $N_{c,a,t}$ denotes the number of octogenarian cases of crime modus operandi $c \in C$ in the location $t \in T$. We evaluate how well the model estimates the number of crimes at each location by the sufficiency ratio $suff_{c,t}$ below.

$$suff_{c,t} = \frac{1}{|A|}\sum_{a \in A} min\left(1.0, \sum_{i \in E_{a,t}} \frac{r_{i,a} \cdot \hat{n}_{c,i}}{N_{c,a,t}}\right) \tag{4}$$

The sufficiency ratio ranges from 0.0 to 1.0, with values closer to 1.0 indicating that the estimation more closely approximates the real data. However, it is important to note that the sufficiency ratio alone does not provide definitive evidence of high accuracy, as it evaluates only the lower limit of the number of crimes. Nevertheless, as shown in Table 3, the fulfillment rate exceeds 80% for

all crime types and areas, suggesting a reasonable level of correspondence between the estimations and the actual data.

Table 3

Fulfillment rates by location of occurrence for each final model.

Modus operandi	Area	Location	N	Mean±Std. of cr
TB	Center	Detached house	73	0.958±0.174
		Condominiums with more than four floors	343	0.891±0.242
		Other houses	41	0.976±0.110
	Surroundings	Detached house	672	0.992±0.046
		Condominiums with more than four floors	856	0.790±0.283
		Other houses	309	0.994±0.068
TC	Center	Detached house	14	0.957±0.159
		Condominiums with more than four floors	50	0.912±0.220
		Other houses	2	1.000±0.000
	Surroundings	Detached house	111	0.974±0.117
		Condominiums with more than four floors	158	0.838±0.243
		Other houses	22	0.991±0.031

6.4. Contribution of explanatory variables

Because the GCN and GBR models are nonlinear, it is not possible to determine the contribution of explanatory variables directly from the data, as is done in multiple regression analysis. Instead, the Integrated Gradients (IG) method is used, which shares mathematical properties such as additivity with SHAP (a widely used explanatory method in machine learning). Consequently, SHAP visualization methods can be applied to interpret the results. Figure 7 presents a summary plot for the GCN model applied to TB in the central area. This plot illustrates the contribution of each variable to the predicted number of crimes for each street, with + indicating a positive contribution and - a negative contribution. Red indicates a high value of the variable, and blue indicates a low value. The value in parentheses to the right of the variable name is the average of the absolute IG values (contribution), and the variables are listed in descending order. In the case of TB in the center area, the top five variables with the largest contribution were bicycles at night, detached houses, street width, apartment buildings, and pedestrians in the daytime. Compared to these results, the same results for the GBR are shown in Figure 8. In this case, the top five variables in order of contribution were bicycles at night, service and entertainment facilities, retail facilities, pedestrians at night, and restaurants. The results are different except for the nighttime pedestrians. In particular, there is a big difference in the point that the street width is ranked high in the GCN model.

The results for additional explanatory variables are presented in Figures A2–A7 in the supplemental material. For the GCN model, the importance of attributes varies depending on the modus operandi and the area being analyzed. However, a common finding is that nighttime bicycle theft consistently shows a high contribution, with a strong positive correlation to the occurrence of crime in all cases. In the central area, street width is a significant variable, ranking highly in importance, and it is ranked sixth in the surrounding area. Interestingly, street width shows a consistent negative correlation with crime occurrences across all areas, indicating that narrower streets may be associated with higher crime rates. Additionally, detached houses make a significant contribution to the prediction of TB regardless of the area, while street length emerges as an important variable with a positive correlation to crime occurrences For car break-ins.

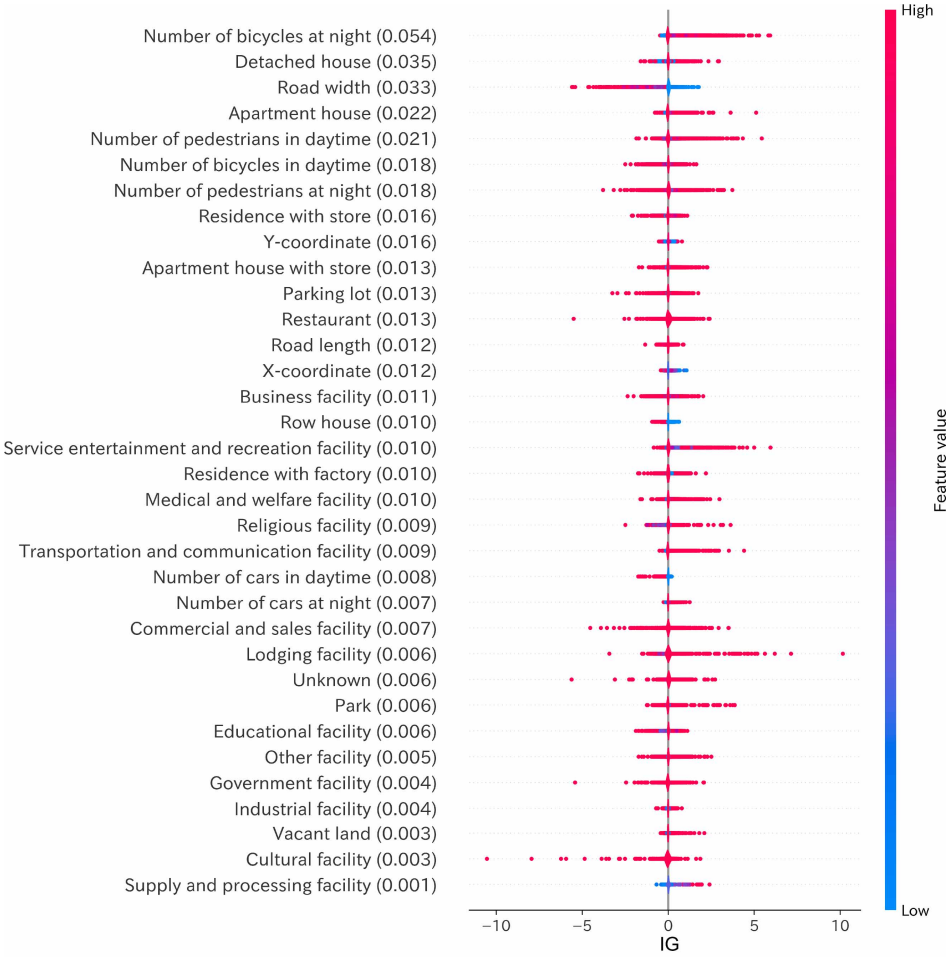

Fig. 7. IG results of GCN model of TB in the central area.

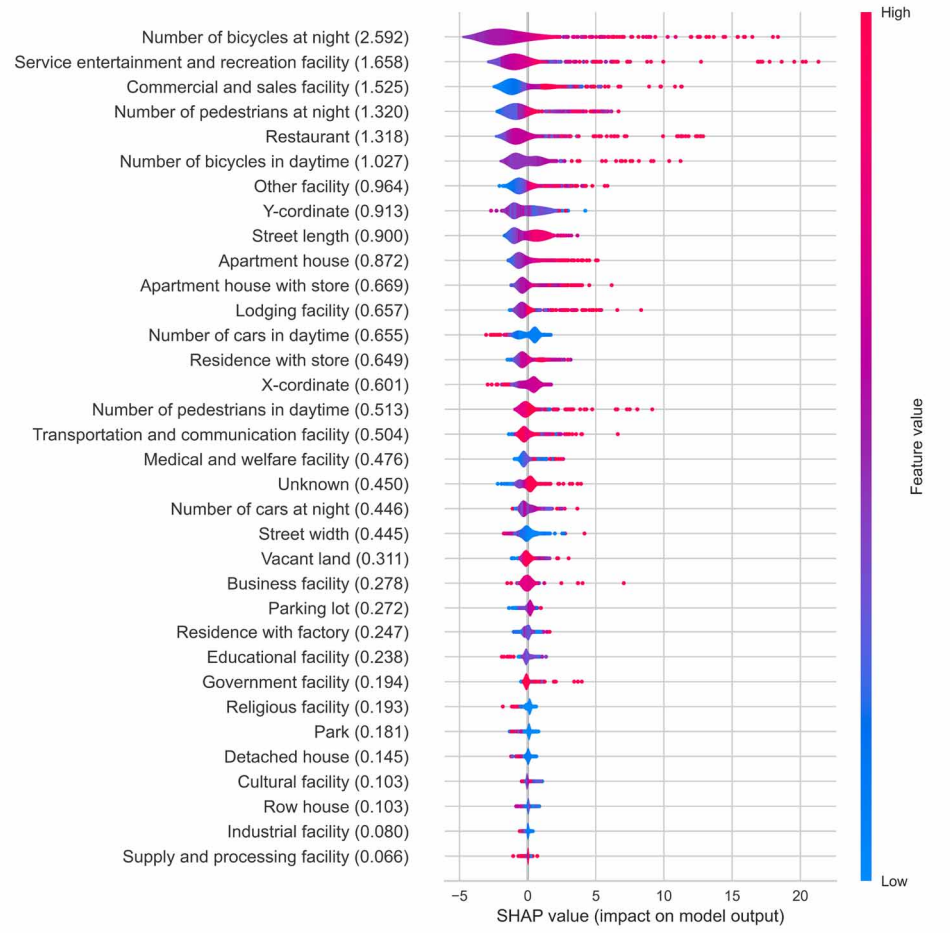

Fig. 8. SHAP results of GBR model for TB in the central area.

In the GBR model, the order of variable contributions differs from that of the GCN model across all cases. Specifically, variables such as service and entertainment facilities, commercial facilities, and restaurants show relatively higher contributions in the GBR model. Conversely, variables like street width and detached houses, which are highly significant in the GCN model, exhibit lower contributions in the GBR model. Despite these differences, the importance of variables related to pedestrian traffic remains consistent across both models.

6.5. Visualization of street level estimates of crime

The proposed model provides estimated crime occurrence values for each street. Figure 9 presents the street level estimated values for TB and TC in the final model. For TB, crime occurrences are more concentrated in the Kita and Minami areas at the city center, as well as around major stations. In contrast, TC exhibits a different pattern, with occurrences more prevalent in the central Minami area and showing a continuous pattern along streets in surrounding areas. This spatial distribution for TC differs from that of TB, which tends to cluster around station vicinities. These differences in spatial patterns, particularly the station-centered clustering of TB versus the linear patterns of TC along peripheral streets, were first identified through the proposed model's ability to estimate crime occurrences at the street level. Such granularity enables a more nuanced understanding of crime dynamics.

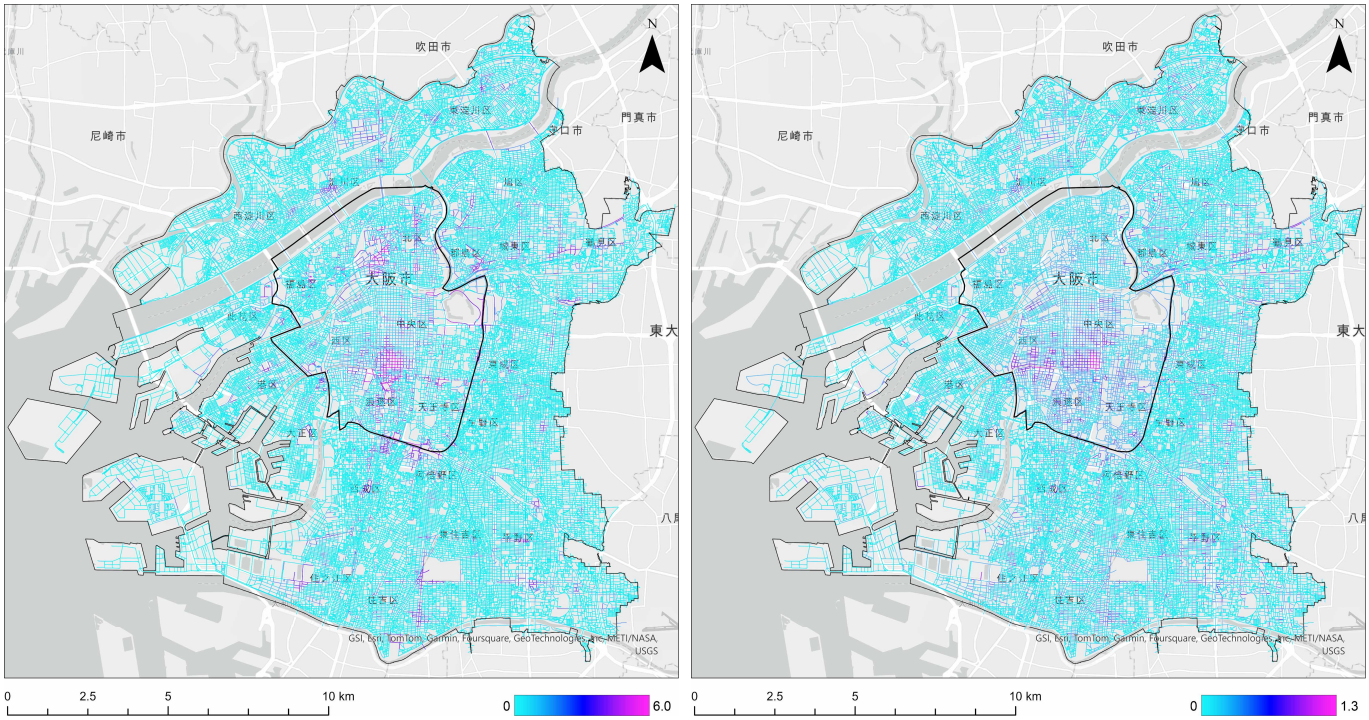

Fig. 9. Estimated number of incidents of TB (left) and TC (right) by the final GCN models for each street.

7. Discussion

From the results obtained, let us discuss the following three points.

7.1. Main factors affecting crime occurrence

Natural surveillance plays a critical role in the design of environments that prevent crime and can be broadly categorized into spatial surveillance and passing vigilance by pedestrians and other passersby. this study incorporates both attributes in the analysis, with explanatory variables constructed to account for these factors. The findings highlight that spatial surveillance is strongly linked to street width and parking areas while passing vigilance has the highest contribution in most patterns. Notably, narrower streets are

associated with higher crime occurrences, emphasizing the significance of poor visibility in crime-prone areas. These results suggest that the visibility and openness of space, as influenced by street width, are important determinants of crime rates. Given the high explanatory power of street width, the study points to a potential direction for further analysis. Future research could involve indexing variables such as the length of the line of sight and the amount of openings in surrounding buildings to provide a more detailed understanding of how spatial characteristics influence crime dynamics.

As for the human flow, the contribution of the number of bicycles is very high, and the number of pedestrians is also high. Since their contribution is positive, it can be interpreted that the high number of people is a factor of criminal opportunities rather than reinforcing surveillance. As described in Section 2, there have been some studies (Andersen 2011, He et al. 2020) that correlate high human presence with crime occurrence, and the trends are similar to those results. However, the reason why the number of bicycles has a larger effect than the number of pedestrians is unclear for TC. Bicycles are widely used in Japan as a means of daily transportation within the city, but they are not as well studied as walkability. It would be possible to conduct research in the direction of analyzing bicycle movement on Japanese streets.

7.2. Significance of analyzing at the street level

The proposed model estimates the number of crimes per street, even though each subregion contains an average of 28 streets in central areas and 55 streets in surrounding areas. This raised concerns that summing the number of crimes per street to estimate the number of crimes per subregion could lead to error accumulation and reduced estimation accuracy. However, cross-validation results showed no statistically significant difference between the two models, demonstrating that the GCN model is a practical and reliable method for crime estimation.

Next, the basis of estimation for the two models, which operate on different scales, can be summarized as follows: The GCN model has higher explanatory power for indicators related to spatial openness, such as parking and street width, while the GBR model has less explanatory power for street width, in particular. For both types of crimes analyzed, the findings contradict Jacobs' assertion that increased human presence reduces crime. Instead, the results suggest that a higher volume of human flow creates more opportunities for crime and increases the number of crime attempts. These findings are consistent with prior research, such as those by Andersen (2011) and He et al. (2020), further supporting the role of human flow as a critical factor in urban crime dynamics.

7.3. Limitations of the study

A key limitation of this study is that the accuracy of the estimated number of crimes per street has not been fully validated. The study originated from the challenge of obtaining precise point-based crime data in Japan. While the author has previously accessed such data from police stations in certain prefectures, the process of acquiring it has proven difficult and inconsistent. The choice of Osaka City as the target area was based on several fundamental conditions relevant to the study. However, the Osaka Prefectural Police, the primary source of the data used, declined to provide detailed crime location data. Consequently, the study relied on aggregated data, which may affect the model's precision at the street level. To enhance the empirical validity of the proposed model, future research should aim to validate the findings using point-based crime occurrence data from other prefectures where such data is accessible. Addressing this limitation is critical for strengthening the model's generalizability and reliability.

8. Concluding remarks

This study proposed a method for analyzing the relationship between street crime and spatial factors at the granular level of street segments, using open crime data aggregated at the census tract level. the proposed method was applied to data on TB and TC

in Osaka City and was compared to a machine learning model based on subregional units. Results demonstrated that the proposed method could statistically estimate crime occurrences with comparable accuracy, despite relying on aggregated data from multiple streets. The model achieved a goodness-of-fit above 0.8 with actual data for the estimated locations, though further validation with point-based crime occurrence data remains necessary. The findings highlighted the importance of micro-spatial attributes, such as street width, in influencing crime dynamics, suggesting the potential applicability of the model for CPTED evaluations. Furthermore, the analysis revealed that human flow, particularly bicycle traffic, significantly contributes to higher crime rates. These insights point to a novel research direction exploring the relationship between bicycle mobility and urban crime in Japanese cities.

Acknowledgements

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by a JSPS Grant-in-Aid for Scientific Research (B) (grant number: 23K28039) and the Nikken Sekkei Research Institute.

References

Agoop Inc. (2019). Point-based flow population data, https://agoop.co.jp/service/dynamic-population-data/

Ali, M. (2020). PyCaret: An open source, low-code machine learning library in Python, https://www.pycaret.org

Andresen, M. A. (2011). The ambient population and crime analysis. The Professional Geographer, 63(2), 193-212.

Andresen, M.A. (2014). Environmental Criminology: Evolution, Theory, and Practice (1st ed.). Routledge.

Anselin, L., Cohen, J., Cook, D., Gorr, W., & Tita, G. (2000). Spatial Analyses of Crime. Criminal Justice 2000, 4, Measurement and Analysis of Crime and Justice. US Department of Justice, Office of Justice Programs, National Institute of Justice, 185541.

Brantingham, P. L., & Brantingham, P. J. (1993). Nodes, paths and edges: Considerations on the complexity of crime and the physical environment, Journal of Environmental Psychology, 13(1), 3-28.

Chicago Data Portal (2012), https://data.cityofchicago.org/

Cohen, B. (1980). Deviant street networks. Lexington, Massachusetts: Lexington Books.

Cohen, L. E., & Marcus F. (1979). Social change and crime rate trends: A routine activity approach. American Sociological Review, 44(4), 588-608.

Cornish, D. B., & Clarke, R. V. eds. (1986). The Reasoning Criminal: Rational Choice Perspectives on Offending. New York: Springer-Verlag.

Curry, G. D., & Irving Spergel, I. A. (1988). Gang homicide, delinquency, and community. Criminology, 26(3), 381-405.

DATA.PLICE.UK. (2013). https://data.police.uk/

De Nadai, M., Xu, Y., Letouzé, E., González, M. C. & Lepri, B. (2020). Socio-economic, built environment, and mobility conditions associated with crime: a study of multiple cities. Sci Rep 10, 13871.

Deisman, W. (2003). CCTV: Literature Review and Bibliography.

Deng, Y., He, R. & Liu, Y. (2023). Crime risk prediction incorporating geographical spatiotemporal dependency into machine learning models, Information Sciences, 646, 119414.

Friedman, J. (2001). Greedy function approximation: A gradient boosting machine, The Annals of Statistics, 29(5), 1189-1232.

Graif, C., & Sampson, R. J. (2009). Spatial heterogeneity in the effects of immigration and diversity on neighborhood homicide rates. Homicide Studies, 13(3), 242-260.

Gu, H., Sui, J., & Chen, P. (2024). Graph representation learning for street-Level Crime Prediction. ISPRS International Journal of Geo-Information. 2024; 13(7), 229.

Hashimoto, S., Yada, A., Kudo, H., Uno, H., & Higuchi, T. (2023). A study on the construction of estimate models for bicycle theft frequent area, JSTE Journal of Traffic Engineering, 9(4), A_18-A_26.
He, L., Páez, A., Jiao, J., An, P., Lu, C., Mao, W., & Long, D. (2020). Ambient population and larceny-theft: A spatial analysis using mobile phone data. ISPRS International Journal of Geo-Information, 2020, 9(6), 342.
He, Q., & Li, J. (2021). The roles of built environment and social disadvantage on the geography of property crime. Cities, 103471, 1-14.
Hillier, B. & Hanson, J. (1984). The Social Logic of Space, Cambridge University Press.
Hipp, J. R., Lee, S., Ki, D., & Kim. J. H. (2022). Measuring the built environment with Google street view and machine learning: Consequences for crime on street segments. J Quant Criminol, 38, 537-565.
Ishikawa, A., Nabeshima, M., & Suzuki, H. (2009). Relationships between the occurrence of snatch and road spatial factors, Journal of Environmental Engineering, 74(635), 55-61.
Jacobs, J. (1961). The Death and Life of Great American Cities, Random House, New York.
Kim, S., & Lee, S. (2023). Nonlinear relationships and interaction effects of an urban environment on crime incidence: Application of urban big data and an interpretable machine learning method, Sustainable Cities and Society, 91, 104419.
Kim, Y. A., & Hipp, J. R. (2018). Physical boundaries and city boundaries: Consequences for crime patterns on street segments? Crime & Delinquency, 64(2), 227-254.
Kishita, H., Shibata, H., Ishibashi, T., Amemiya, M., & Hino, K. (2016). A study on crime situation and the prevention measures for convenience store, Journal of the City Planning Institute of Japan, 51(3), 350-356.
Lee, I., Jung, S., Lee, J., & Macdonald, E. (2019). Street crime prediction model based on the physical characteristics of a streetscape: Analysis of streets in low-rise housing areas in South Korea. Environment and Planning B: Urban Analytics and City Science, 46(5), 862-879.
Lee, N., & Contreras, C. (2021). Neighborhood walkability and crime: Does the relationship vary by crime type? Environment and Behavior, 53(7), 753-786.
Liu, L., Jiang, H., and He, P., Chen, W., Liu, X., Gao, J., & Han, J. (2020). On the variance of the adaptive learning rate and beyond, Proceedings of the Eighth International Conference on Learning Representations (ICLR 2020).
Lundberg, S. M. & Lee, S. I. (2017). A unified approach to interpreting model predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). 4768-4777.
Marchment, Z., Frith, M. J., Morrison, J., & Gill, P. (2021). A multi-level analysis of risky streets and neighbourhoods for dissident republican violence in Belfast. ISPRS International Journal of Geo-Information, 10(11), 765.
Maruthaveeran, S., & Bosh, C. K. (2015). Fear of crime in urban parks - What the residents of Kuala Lumpur have to say ?, Urban Forestry & Urban Greening 14(3), 702-713.
Ministry of Land, Infrastructure, Transport and Tourism, Japan. (2020). PLATEAU, https://www.mlit.go.jp/plateau/
National Police Agency (2018), Crime Open Data Links, https://www.npa.go.jp/toukei/seianki/hanzaiopendatalink.html
Newman O. (1972). Defensible Space: Crime Prevention Through Urban Design. New York: Macmillan.
NYC Open Data (2013). https://opendata.cityofnewyork.us/
Osaka Prefecture Police (2023). Crime Open Data Site, https://www.police.pref.osaka.lg.jp/seikatsu/9290.html
Osborn, D., Trickett, A., & Elder, R. (1992). Area characteristics and regional variates as determinants of area property crime levels. Journal of quantitative criminology, 8(3), 265-285.
PyG. (2018). https://pytorch-geometric.readthedocs.io/en/latest/index.html#
PyTorch. (2016). https://pytorch.org/
Roncek, D. W., & Maier, P. A. (1991). Bars, blocks, and crimes revisited: Linking the theory of routine activities to the empiricism of "hot spots". Criminology, 29, 725-755.

Sampson, R. J., Raudenbush, S. W. & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. Science, 277, 918-924.
Sampson, R., & Groves, W. (1989). Community structure and crime: Testing socialdisorganization theory. American Journal of Sociology, 94(4), 774-802.
Saraiva, M., & Teixeira, B. (2023). Exploring the spatial relationship between street crime events and the distribution of urban greenspace: The case of Porto, Portugal. ISPRS International Journal of Geo-Information, 12(12), 492.
Shaw, C. R., & McKay, H. D. (1942). Juvenile Delinquency and Urban Areas. University of Chicago Press.
Statistics Bureau of Japan (2021). 2020 Population Census, https://www.stat.go.jp/english/data/kokusei/index.html
Summers, L., & Johnson, S. D. (2017). Does the configuration of the street network influence where outdoor serious violence takes place? Using space syntax to test crime pattern theory. J Quant Criminol, 33, 397-420.
Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. In Proc. 34th Int. Conf. Mach. Learn. 70, 3319-3328.
Sypion-Dutkowska, N., & Leitner, M. (2017). Land use influencing the spatial distribution of urban crime: A case study of Szczecin, Poland. ISPRS International Journal of Geo-Information, 6(3), 74.
Takagi, N., & Takizawa. A. (2021). A study on the method for estimating walking data by use of point based GPS data, Proceedings of the 44th Symposium on Computer Technology of Information, Systems and Applications, 339-342.
Takizawa, A. (2013). Emerging pattern based street crime analysis, Journal of Architecture and Planning, 78(686), 957-967.
Toronto Police Service, (2022). https://data.torontopolice.on.ca/pages/open-data
United Nations Office on Drugs and Crime (2020). dataUNODC, https://dataunodc.un.org/
Welsh, B.C. & Farrington, D.P. (2008). Effects of improved street lighting on crime. Campbell Systematic Reviews, 4, 1-51.
Wolfe, M. K., & Mennis, J. (2012). Does vegetation encourage or suppress urban crime? Evidence from Philadelphia, PA. Landscape and Urban Planning, 108(2-4), 112-122.
Xavier, B., & Thomas, L. (2017). Residual gated graph ConvNets, arXiv preprint, arXiv:1711.07553.
Xiang, L., Sheng J. & Liao, P. (2023). Understanding the relationship between the spatial configuration and the crime rate of downtown eastside in Vancouver, Canada, Habitat International, 137, 102847.
Xie, H., Liu, L., Yue, H. (2022). Modeling the effect of streetscape environment on crime using street view images and interpretable machine-learning technique. International Journal of Environmental Research and Public Health. 2022; 19(21), 13833.
Zhang, Y., & Cheng, T. (2020). Graph deep learning model for network-based predictive hotspot mapping of sparse spatio-temporal events, Computers, Environment and Urban Systems, 79, 101403.

Supplementary materials

Appendix A: Additional tables and figures

Table A1

Basic statistics of explainable variables (N = 90,728).

Variable	Mean	SD	CV	IQR	Min	Median	Max
Detached house (m^2)	370.3	608.8	1.6	508.6	0	110.1	15,189.3
Row house (m^2)	65.7	192.9	2.9	0	0	0	4,226.8
Apartment house (m^2)	673.5	3,098.9	4.6	85.9	0	0	132,969
Apartment house with store (m^2)	222.6	2,885.9	13	0	0	0	716,015.7
Residence with store (m^2)	84.6	227.3	2.7	0	0	0	17,229.1
Residence with factory (m^2)	22.4	126.8	5.7	0	0	0	7,993.9
Commercial and sales facility (m^2)	94.6	3,567.3	37.7	0	0	0	587,633.1
Restaurant (m^2)	18.4	262.2	14.2	0	0	0	32,084.1
Business facility (m^2)	351.7	3,833.3	10.9	0	0	0	516,561.5
Service entertainment and recreation facility (m^2)	37.9	1,205.1	31.8	0	0	0	285,649.1
Educational facility (m^2)	82.8	859.2	10.4	0	0	0	74,796.9
Cultural facility (m^2)	15.7	728.7	46.4	0	0	0	153,840.5
Religious facility (m^2)	16.9	213.8	12.6	0	0	0	24,071.4
Medical and welfare facility (m^2)	55.6	1,143.8	20.6	0	0	0	154,058.7
Lodging facility (m^2)	57.6	2,635.5	45.7	0	0	0	417,905.4
Industrial facility (m^2)	136.4	1,489.5	10.9	0	0	0	190,011
Supply and processing facility (m^2)	17.7	848.7	47.9	0	0	0	146,391.4
Transportation and communication facility (m^2)	154.7	2,325.3	15	0	0	0	196,171.8
Government facility (m^2)	24.6	810.7	33	0	0	0	115,143.4
Other facility (m^2)	21.3	345.7	16.3	0	0	0	33,432.7
Unknown (m^2)	318.4	3,822.6	12	115.3	0	0	671,355.4
Parking lot (m^2)	176.1	758.8	4.3	103.3	0	0	51,278.9
Park (m^2)	753	9,071.7	12	0	0	0	584,261.7
Vacant land (m^2)	583.4	6,123.6	10.5	0	0	0	289,056.4
Street width (m)	7.6	6.2	0.8	4.4	0.2	5.9	78.4
Street length (m)	56.7	49.9	0.9	45.1	0.5	46.3	1,991.0
X-cordinate (m, Cartesian coordinate system)	-44,640.8	3,377.4	-	5,191.3	-55,963.2	-44,417.2	-36,714.5
Y-cordinate (m, Cartesian coordinate system)	-147,346.5	4,784.9	-	8,012.6	-156,610.8	-147,497.9	-136,510.1
Number of pedestrians in daytime	45.1	212.5	4.7	21.2	0	4.6	12,170.2
Number of bicycles in daytime	9.8	28.2	2.9	6.8	0	1.4	1,084.0
Number of cars in daytime	11.5	62.3	5.4	3.8	0	0.4	4,759.2
Number of pedestrians at night	24.6	127.6	5.2	11.2	0	2	7,959.6
Number of bicycles at night	4.6	13.9	3.1	3	0	0.8	463.2
Number of cars at night	4.1	22.2	5.4	1.4	0	0	1,635.8

Table A2

Coalition of variables (N = 1,904).

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19	A20	A21	A22	A23	A24	A25	A26	A27	A28	A29	A30	A31	A32	A33	A34	A35
A1 Detached house																																			
A2 Row house	0.58																																		
A3 Apartment house	0.03	-0.04																																	
A4 Apartment house with store	-0.08	-0.05	0.21																																
A5 Residence with store	0.50	0.47	0.05	0.11																															
A6 Residence with factory	0.27	0.23	-0.01	0.01	0.29																														
A7 Commercial and sales facility	-0.07	-0.05	0.03	0.03	-0.03	-0.04																													
A8 Restaurant	-0.10	-0.07	-0.04	0.02	0.02	-0.06	0.06																												
A9 Business facility	-0.24	-0.16	-0.01	0.12	-0.12	-0.11	0.05	0.08																											
A10 Service entertainment and recreation facility	-0.08	-0.05	-0.01	0.07	-0.01	-0.03	0.04	0.16	0.02																										
A11 Educational facility	0.00	-0.02	0.10	0.05	-0.02	0.00	0.01	-0.02	0.03	0.00																									
A12 Cultural facility	-0.08	-0.05	0.03	0.09	-0.06	-0.04	0.01	0.00	0.40	0.01	0.06																								
A13 Religious facility	0.07	0.05	-0.04	0.02	0.07	0.02	-0.02	0.00	0.01	-0.02	0.07	0.01																							
A14 Medical and welfare facility	-0.03	-0.03	0.07	0.08	-0.01	0.00	0.01	-0.01	0.03	-0.01	0.10	0.08	-0.02																						
A15 Lodging facility	-0.13	-0.08	0.00	0.08	-0.08	-0.07	0.10	0.09	0.30	0.04	0.04	0.21	-0.01	0.07																					
A16 Industrial facility	-0.06	-0.07	-0.05	-0.07	-0.11	0.11	-0.02	-0.04	-0.05	0.03	-0.05	-0.02	-0.06	-0.03	-0.04																				
A17 Supply and processing facility	-0.05	-0.05	-0.03	-0.04	-0.09	-0.03	-0.01	-0.02	0.02	0.02	-0.01	0.04	-0.03	0.07	-0.01	0.06																			
A18 Transportation and communication facility	-0.16	-0.12	-0.06	-0.04	-0.14	-0.06	0.04	-0.02	0.03	-0.01	-0.04	0.04	-0.06	-0.03	0.03	0.26	0.19																		
A19 Government facility	-0.10	-0.06	0.04	0.03	-0.06	-0.05	0.01	0.01	0.08	0.02	0.03	0.03	0.02	0.15	0.03	-0.02	0.00	0.03																	
A20 Other facility	-0.04	-0.01	0.11	0.07	0.04	0.05	0.03	0.08	0.06	0.08	-0.03	0.02	-0.01	-0.02	0.15	0.00	-0.03	-0.01	-0.02																
A21 Unknown	-0.11	-0.07	0.06	0.04	-0.04	-0.06	0.26	0.04	0.26	0.25	0.09	0.04	-0.01	0.05	0.19	0.10	0.00	0.15	0.05	0.08															
A22 Parking lot	0.22	0.08	0.14	0.03	0.07	0.10	0.00	-0.04	0.08	-0.03	0.03	0.25	-0.02	0.01	0.04	0.19	0.08	0.31	-0.05	0.00	0.07														
A23 Park	-0.04	-0.05	0.02	-0.03	-0.09	-0.05	-0.02	-0.01	0.00	0.02	0.04	0.34	0.02	0.00	0.00	0.02	0.03	0.05	0.01	0.02	0.07	0.07													
A24 Vacant land	0.03	-0.05	0.04	-0.02	-0.10	-0.05	0.05	-0.03	0.09	0.00	0.01	0.20	-0.04	0.02	0.03	0.21	0.19	0.30	-0.02	0.03	0.12	0.26	0.15												
A25 Street width	-0.52	-0.40	-0.07	0.06	-0.32	-0.19	0.11	0.03	0.25	0.06	-0.01	0.13	-0.02	0.02	0.15	0.07	0.09	0.32	0.10	0.03	0.17	-0.04	0.10	0.11											
A26 Street length	0.61	0.40	0.37	0.07	0.39	0.24	0.02	-0.02	-0.03	0.00	0.12	0.12	0.03	0.04	-0.02	0.15	0.08	0.14	-0.02	0.04	0.07	0.52	0.12	0.30	-0.28										
A27 X-cordinate	0.24	0.15	0.03	0.00	0.13	0.23	-0.03	-0.01	-0.09	-0.05	0.04	-0.07	0.06	0.01	-0.05	-0.17	-0.09	-0.30	-0.03	0.00	-0.10	-0.03	-0.05	-0.24	-0.32	0.03									
A28 Y-cordinate	-0.09	-0.06	0.17	0.08	-0.03	0.02	0.02	0.04	0.13	0.02	0.00	-0.01	0.01	0.03	0.05	0.01	0.01	-0.06	0.03	0.01	0.09	0.02	0.04	0.02	0.02	0.11	-0.02								
A29 Number of pedestrians in daytime	-0.16	-0.10	0.06	0.13	0.01	-0.06	0.44	0.20	0.48	0.11	0.03	0.03	0.01	0.05	0.30	-0.06	-0.02	0.00	0.06	0.04	0.41	-0.02	0.00	-0.03	0.26	0.04	-0.03	0.19							
A30 Number of bicycles in daytime	-0.08	-0.03	0.15	0.25	0.15	-0.01	0.31	0.21	0.37	0.14	0.05	0.12	0.06	0.09	0.26	-0.07	-0.02	-0.02	0.11	0.09	0.30	0.08	0.09	-0.03	0.26	0.23	0.07	0.16	0.71						
A31 Number of cars in daytime	0.07	-0.07	0.12	0.00	-0.03	-0.02	0.02	-0.02	0.00	0.02	0.04	0.12	-0.06	0.03	0.02	0.18	0.12	0.18	0.00	-0.01	0.05	0.34	0.07	0.17	0.14	0.39	-0.07	0.09	0.05	0.20					
A32 Number of pedestrians at night	-0.14	-0.09	0.06	0.11	0.02	-0.06	0.41	0.28	0.41	0.12	0.02	0.01	0.01	0.03	0.29	-0.06	-0.03	0.00	0.04	0.03	0.41	-0.04	-0.01	-0.03	0.22	0.04	-0.03	0.18	0.97	0.67	0.03				
A33 Number of bicycles at night	-0.11	-0.05	0.14	0.25	0.13	-0.05	0.31	0.33	0.38	0.16	0.05	0.10	0.07	0.07	0.28	-0.10	-0.03	-0.04	0.10	0.10	0.31	0.04	0.08	-0.04	0.24	0.18	0.04	0.17	0.76	0.96	0.13	0.75			
A34 Number of cars at night	0.07	-0.06	0.14	0.02	-0.01	-0.03	0.03	0.00	-0.01	0.02	0.04	0.09	-0.05	0.03	0.04	0.15	0.08	0.13	-0.01	0.00	0.06	0.30	0.06	0.14	0.12	0.37	-0.04	0.11	0.09	0.24	0.99	0.08	0.19		
A35 TB	-0.04	0.00	0.26	0.32	0.26	-0.03	0.22	0.36	0.17	0.22	0.07	0.03	0.00	0.06	0.20	-0.11	-0.05	-0.07	0.04	0.13	0.17	0.07	0.04	-0.06	0.02	0.21	0.05	0.15	0.47	0.59	0.06	0.48	0.65	0.11	
A36 TC	0.07	0.03	0.22	0.28	0.30	0.09	0.08	0.18	0.09	0.09	0.03	0.03	0.06	0.04	0.08	-0.06	-0.06	-0.05	-0.01	0.12	0.06	0.15	0.01	-0.05	-0.05	0.24	0.08	0.01	0.21	0.41	0.04	0.21	0.43	0.07	0.54

Table A3

Locations of TBs from 2018 to 2019.

Location	Number of cases		Ratio	
	Center	Surroundings	Center	Surroundings
Detached houses	96	1,265	0.01	0.08
Apartment buildings under 3 stories etc.	43	394	0.01	0.02
Apartment complexes of 4 or more stories	1,153	3,318	0.14	0.20
Parking lot	2,218	5,316	0.28	**0.32**
On the street	3,907	4,658	**0.49**	0.28
Others	625	1,642	0.08	0.10
Total	8,042	16,593	1.00	1.00

Table A4

Locations of TCs from 2018 to 2019.

Location	Number of cases		Ratio	
	Center	Surroundings	Center	Surroundings
Detached houses	17	125	0.01	0.03
Apartment buildings under 3 stories etc.	2	23	0.00	0.01
Apartment complexes of 4 or more stories	57	201	0.03	0.05
Parking lot	1,077	2,263	**0.53**	**0.54**
On the street	774	1,105	0.38	0.26
Others	96	475	0.05	0.11
Total	2,023	4,192	1.00	1.00

Table A5

Average accuracy of 10-fold cross-validations of GCN models.

Modus operandi	Area	d	λ	Epoch	R2	RMSE
TB	Center	0	1.0e-4	78	0.669±0.134	9.685±1.718
		1	↑	**237**	**0.700±0.130**	**9.183±1.963**
		2	↑	30	0.671±0.143	9.597±1.624
		3	↑	23	0.647±0.113	10.15±1.980
	Surroundings	0	1.0e-4	135	0.497±0.165	9.390±2.019
		1	↑	**218**	**0.550±0.145**	**8.891±1.921**
		2	↑	85	0.511±0.137	9.261±1.777
		3	↑	185	0.441±0.197	9.831±1.889
TC	Center	0	3.3e-5	196	0.460±0.141	3.106±0.427
		1	↑	58	0.479±0.149	3.034±0.384
		2	↑	**24**	**0.522±0.171**	**2.887±0.384**
		3	↑	24	0.459±0.224	3.038±0.471
	Surroundings	**0**	**1.0e-4**	**45**	**0.317±0.097**	**2.328±0.143**
		1	↑	36	0.268±0.125	2.409±0.174
		2	↑	25	0.281±0.114	2.387±0.175
		3	↑	13	-0.215±0.839	2.913±0.759

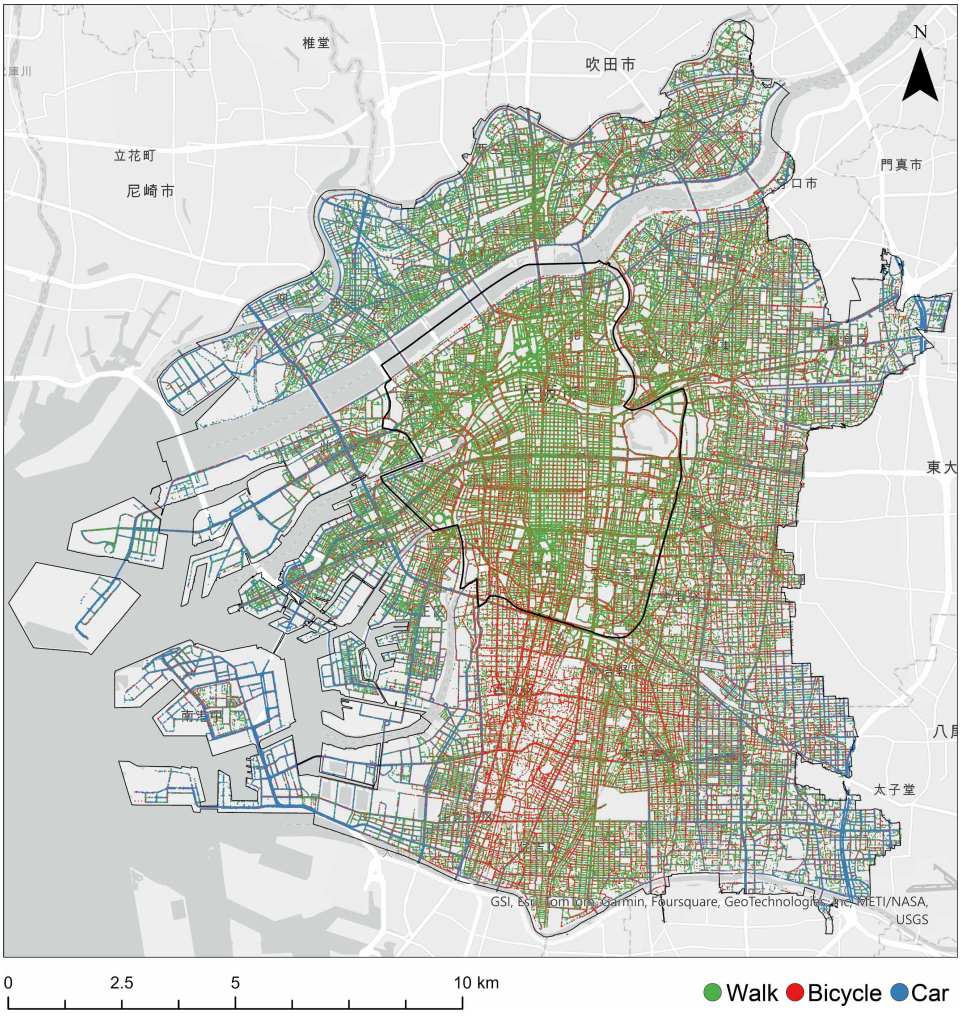

Fig. A1. Distribution of human flow points (green: pedestrians, red: bicycles, blue: automobiles).

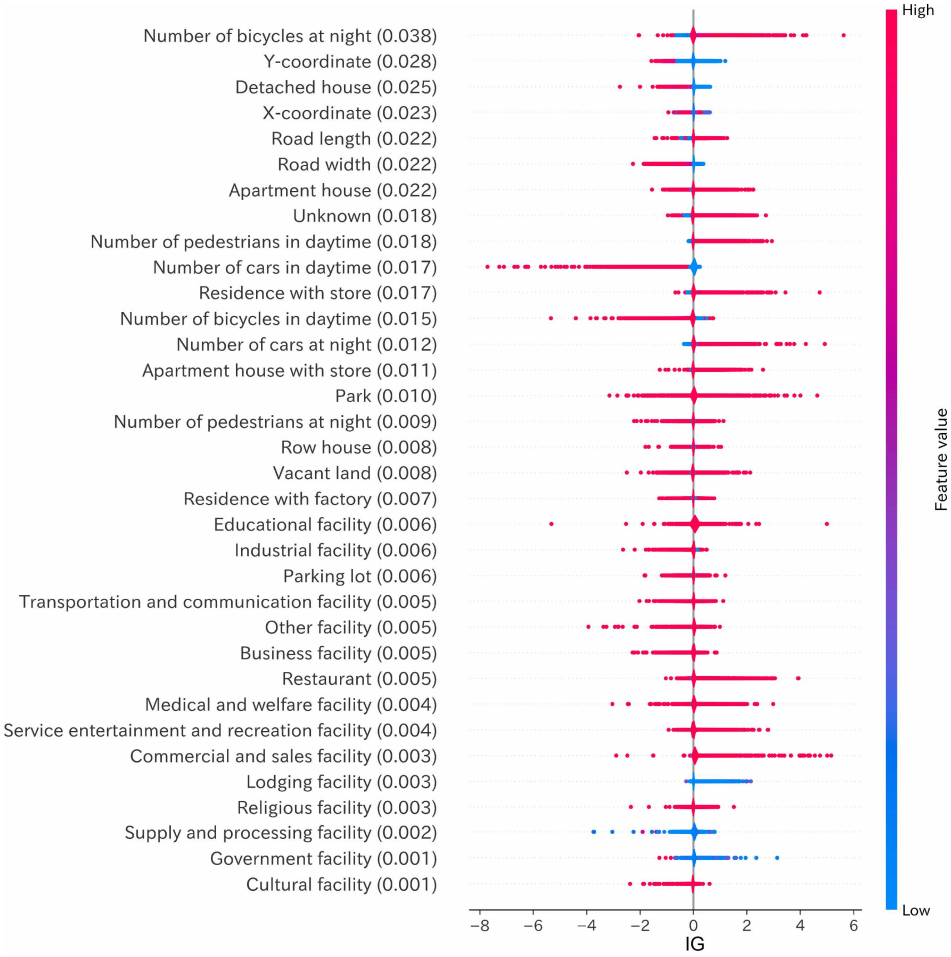

Fig. A2. IG results of GCN model of TB in surrounding area.

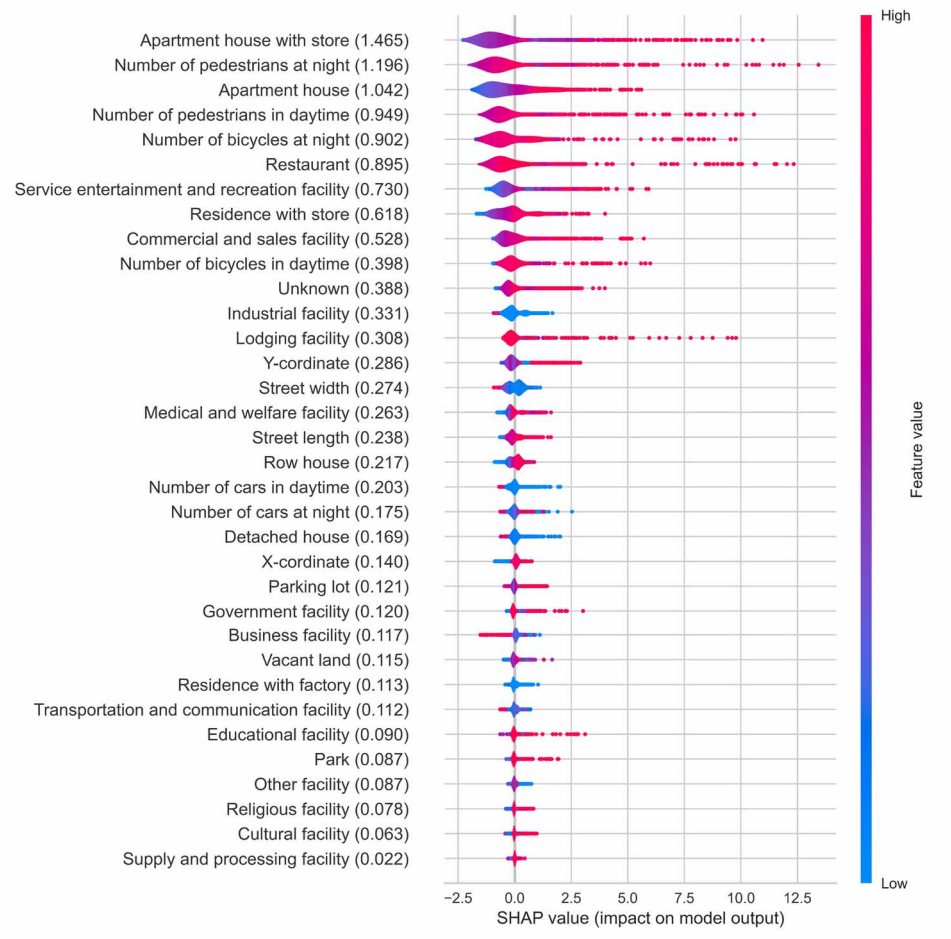

Fig. A3. SHAP results of GBR model for TB in surrounding area.

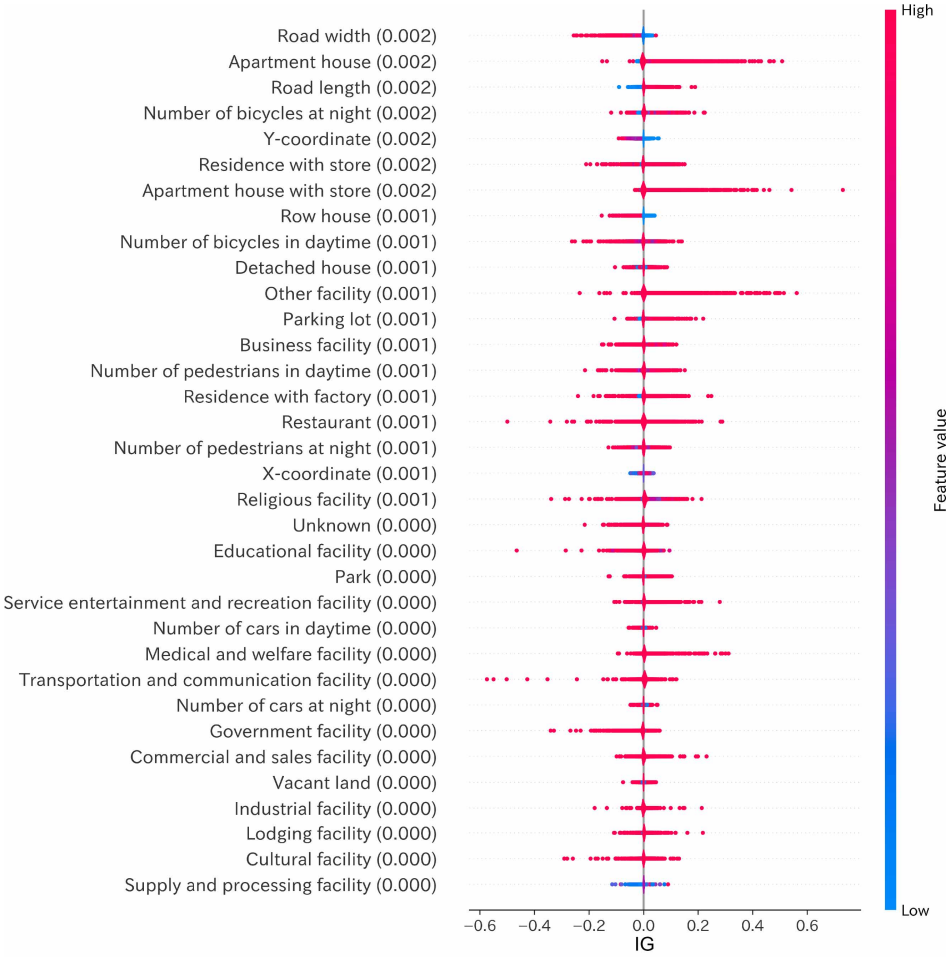

Fig. A4. IG results of GCN model of TC in central area.

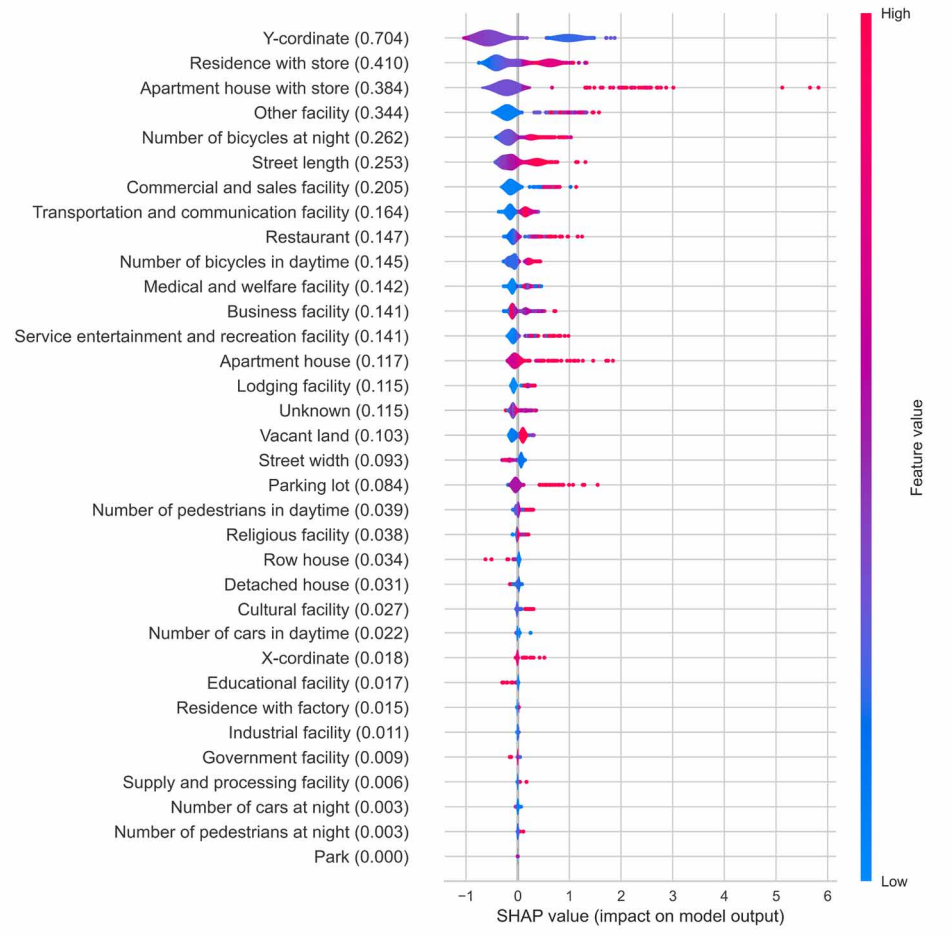

Fig. A5. SHAP results of GBR model for TC in central area.

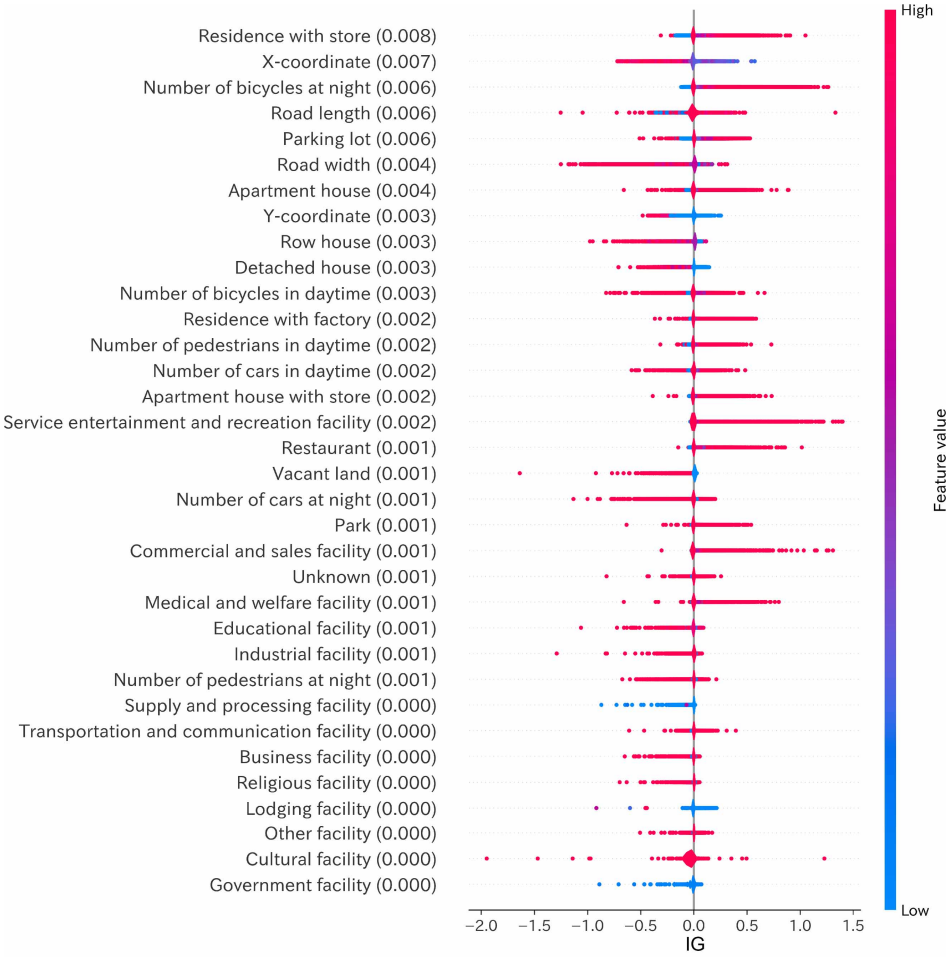

Fig. A6. IG results of GCN model of TC in surrounding area.

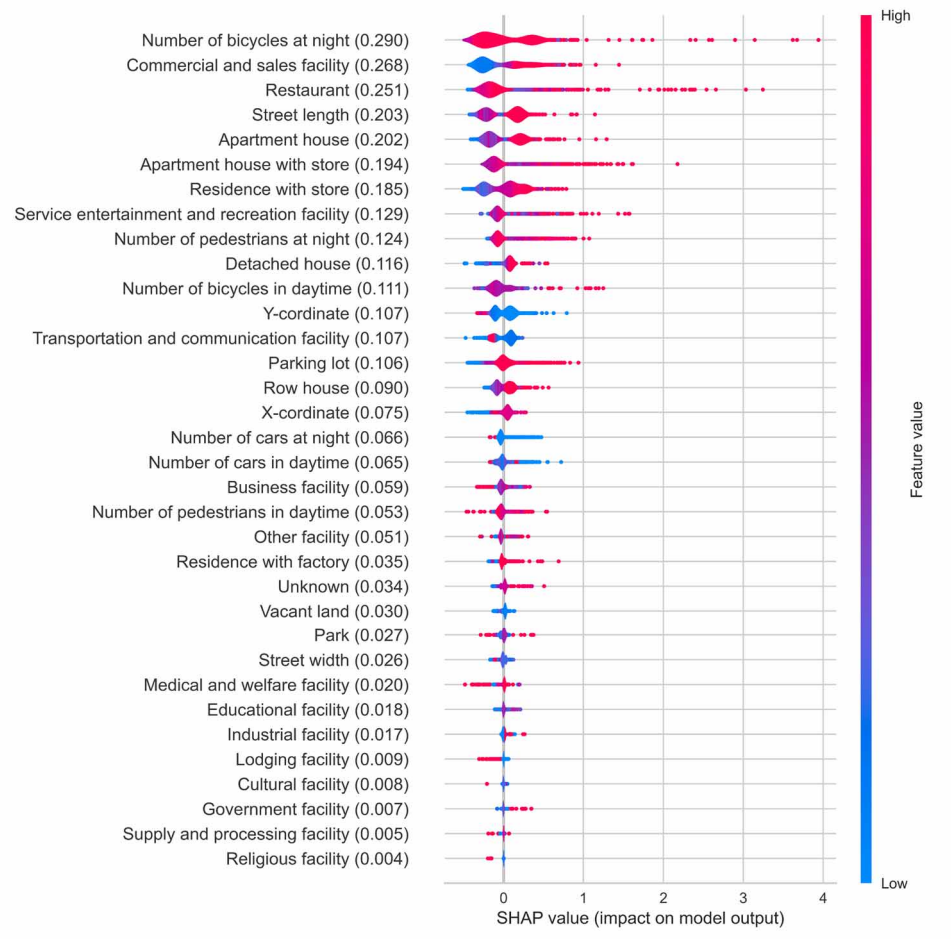

Fig. A7. SHAP results of GBR model for TC in surrounding area.

Appendix B. How to estimate the length and width of each street

First, let a and p be the area and perimeter length of each street polygon, respectively. Next, let's assume that the area of the street is a rectangle, and let's call the shorter side of the rectangle the street width and the longer side the street length and denote them as w and $l(0 < w \leq l)$. Then, the area and the perimeter length of the rectangle can be written as $w \cdot l$ and $2(w + l)$, respectively. In order to approximate the original area with a rectangle, we solved the linear programming problem, which minimizes the sum of the absolute values of the differences in area and perimeter length of the area defined in equation (B1).

$$|w \cdot l - a| + |2(w + l) - p| \tag{B1}$$