Taylor & Francis
Taylor & Francis Group

Journal of Asian Architecture and Building Engineering

ISSN: 1346-7581 (Print) 1347-2852 (Online) Journal homepage: www.tandfonline.com/journals/tabe20

Correlation analysis between urban environment features and crime occurrence based on explainable artificial intelligence techniques

Geunhan Kim, Youngtae Cho, Ju-Heon Lee & Gunwon Lee

To cite this article: Geunhan Kim, Youngtae Cho, Ju-Heon Lee & Gunwon Lee (29 Oct 2024): Correlation analysis between urban environment features and crime occurrence based on explainable artificial intelligence techniques, Journal of Asian Architecture and Building Engineering, DOI: 10.1080/13467581.2024.2421260

To link to this article: https://doi.org/10.1080/13467581.2024.2421260

Published online: 29 Oct 2024.

JOURNAL OF ASIAN ARCHITECTURE AND BUILDING ENGINEERING
https://doi.org/10.1080/13467581.2024.2421260

Taylor & Francis
Taylor & Francis Group

URBAN PLANNING AND DESIGN

OPEN ACCESS Check for updates

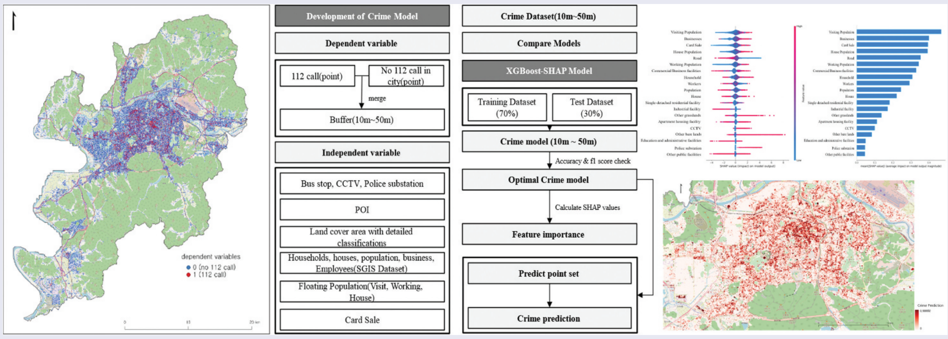
Correlation analysis between urban environment features and crime occurrence based on explainable artificial intelligence techniques

Geunhan Kim[a], Youngtae Cho[b], Ju-Heon Lee[c] and Gunwon Lee[d]

[a]Water and Land Research Group, Korea Environment Institute, Sejong, Republic of Korea; [b]Land & Housing Research Institute, Daejeon, Republic of Korea; [c]Department of Geoinformatics, University of Seoul, Seoul, Republic of Korea; [d]Department of Architecture, Korea University, Seoul, Republic of Korea

ABSTRACT

Crime and urban environments are considered to be closely related. However, there exists no clear understanding of this phenomenon. Therefore, this study analyzes the relationship between various urban environmental variables and crime occurrences and provides insights into the optimal placement of crime prevention facilities by developing a crime prediction map based on a paradigm of the Daegu city. To achieve this, we used 373,387 crime reports from Daegu as dependent variables and 370,000 random points. Independent variables included information such as the point of interest, land use, land cover, floating population, and card sales. The developed crime prediction map created using the model was used to evaluate the adequacy of CCTV installation locations and identify areas requiring new CCTV installations. The performances of various machine-learning models were compared and the XGBoost model (accuracy of 89.7 % and precision of 89.8 %) was selected. Key variables influencing crime report data were identified using the SHAP(SHapley Additive explanation) method. To analyze the spatial explanatory power of the relationship between crime and urban environmental variables, various buffer distances were tested, and a 20 m buffer distance was derived. The results of this study are expected to provide valuable data for crime prevention policies.



ARTICLE HISTORY
Received 29 June 2024
Accepted 21 October 2024

KEYWORDS
Crime occurrence; urban environment feature; closed-circuit television; extreme gradient boosting; SHapley additive explanations

1. Introduction

Crime negatively influences the economy, society, culture, and lives of urban dwellers of a city. Crimes incur social costs and slow down the economic growth of a city (Cullen and Levitt 1999; Detotto and Vannini 2010). An increased crime rate causes psychological fear among urban dwellers. This fear leads to residents' decisions to leave the city or avoid outdoor activities, thus resulting in the economic and sociocultural decline of the city (Schwartz, Susin, and Voicu 2003). Therefore, policymakers and researchers have implemented various measures to prevent crime. Brantingham & Faust (1976) categorized crime prevention measures into three levels of activities, namely primary, secondary, and tertiary levels (Brantingham and Brantingham 1981). First, primary crime prevention involves the identification and elimination of physical and social environmental conditions that trigger criminal behavior or provide any related opportunities. Prevention measures typically include improving the environmental design and enhancing the neighborhood surveillance and private security by installing crime prevention facilities. Secondary crime prevention includes the early identification of potential criminals through the studies on criminal psychology (e.g., psychological state and growth processes of criminals) or criminal behaviors. Tertiary crime prevention is related to increasing criminal penalties for crimes to deter offenders from committing crimes. Primary crime prevention measures are typically preferred over

CONTACT Gunwon Lee upnd.cla@gmail.com Department of Architecture, Korea University, Seoul, Republic of Korea

secondary and tertiary prevention measures because of their pre-emptive nature and effectiveness (Andresen and Jenion 2008). Thus, identifying environmental conditions that are highly correlated with criminal behavior in advance, and creating a physical and social environment for crime prevention are crucial (Johnson and Bowers 2010).

Studies have examined the relationship between crime occurrences, various features of cities (e.g., road systems, urban decay, land use and land cover (LULC), and the spatial structures in cities. The results of a study have indicated that crime rates are higher in the central areas of cities (such as station districts) because of new buildings and a larger floating population (Weisburd et al. 2020). In terms of land use, areas with a high concentration of commercial facilities are typically associated with higher crime rates (He and Li 2021). Additionally, even within the same commercial districts, areas with a high concentration of nighttime businesses have higher crime rates (Twinam 2017). The increased crime rates in a dense residential area could be attributed to a large resident population (Sampson 1983). By contrast, in locations in which the density of commercial facilities and residential areas is high, the crime rate decreases as a function of the density (Browning et al. 2010). Furthermore, complex land uses can also affect crime rates (Yirmibesoglu and Ergun 2007). Among land cover types, green spaces, and vegetation exhibit a complex relationship with crime rates. Green spaces deter criminal activities by separating areas from visitors and encouraging social activities (Lin, Wang, and Huang 2021; Newman 1972; Wolfe and Mennis 2012). By contrast, studies have indicated that green spaces can conceal criminal activities and that trees can increase crime by hindering natural surveillance (Gilstad-Hayden et al. 2015; Kim and Hipp 2018; Maruthaveeran and van den Bosch 2015).

Defining the relationship between urban characteristics, spatial urban structures, and crime occurrence is challenging owing to their complex inter-relationships. Crime occurrence is not driven by a single factor but rather by multiple contributing factors (Cowen, Loiderback, and Roy 2019). The complexity of the relationship between independent and dependent variables has led to the application of nonlinear models rather than traditional linear models (Hipp et al. 2022; Jang and Lee 2022). Specifically, it is necessary to analyze various environmental factors comprehensively and integrate them with spatial information (Giménez-Santana, Medina-Sarmiento, and Miró-Llinares 2018; Thomas, Harris, and Drawve 2022). A representative method for this is risk terrain modeling (RTM). RTM spatially analyzes risk factors to visualize the risk levels of specific areas and predict locations where incidents are likely to occur in the future (Caplan et al. 2015). Many previous studies have analyzed these relationships across various regions using methods like RTM. Studies have predicted crime occurrence in multiple locations, including Chicago (Caplan et al. 2015), Atlanta (Bagwell et al. 2024), Santiago, Chile (Vildosola et al. 2019), Valencia, Spain (Briz-Redón, Mateu, and Montes 2021), and the North Rhine-Westphalia region in Germany (Schwarz and Seidensticker 2023) by considering the unique characteristics of each area. These studies expected that the findings would enable law enforcement authorities to determine where to focus their crime prevention resources and develop strategies accordingly (Caplan et al. 2020; Garnier, Caplan, and Kennedy 2018). Moreover, these studies emphasized that combining RTM with machine-learning models could enhance performance in crime-data-based predictions (Marchment and Gill 2021).

Although crime occurrences have been predicted by considering regional characteristics as presented in these research methods, these approaches have limitations in explaining precise locations from the urban planning and design perspectives. For instance, to evaluate the appropriateness of existing CCTV locations and plan future CCTV installations, it is necessary to derive detailed location-specific crime occurrence probabilities, which is difficult to achieve using these precise location information and crime occurrence probability models.

With the increasing application of artificial intelligence (AI)-based machine learning across various fields, nonlinear models have attracted interest in this area of research. Crime prediction models can incorporate data such as past crime incidents, various environmental features, and time (Zhang et al. 2022). Machine-learning models tend to exhibit superior model fits than traditional linear regression models. Previous studies have demonstrated the outstanding performance and efficiency of machine-learning techniques in predicting crime in urban environments (Lan, Liu, and Eck 2021).

Machine-learning techniques, such as random forest (Alves, Ribeiro, and Rodrigues 2018; Wheeler and Steenbeek 2020), LightGBM (Tong et al. 2021), neural network (NN) models (Rummens, Hardyns, and Pauwels 2017), deep NNs (DNNs) (Kang, Kang, and Choo 2017), long short-term memory (LSTM) (Zhang et al. 2020), and convolutional neural network (CNN) (Wang et al. 2020) models have been extensively applied for crime prediction research. These AI-based techniques can be used to predict the locations where crimes are likely to occur. However, because of the limitations of typical black-box models, interpreting why the predicted locations have high risks is difficult (Wheeler and Steenbeek 2020). Therefore, applying the aforementioned models is not practical in cases wherein the police force wants to establish crime prevention and control strategies (Zhang et al. 2020).

In this study, we used explainable AI (XAI), examined the independent variable with impacts on crime among urban environmental conditions, and predicted the points where crime could occur. XAI can compensate for the limitations of existing black-box models by providing analyzed results so that humans can understand the reasons for these predictions (Kim and Lee 2023; Kim, McCarty, and Jeong 2023; Zhang et al. 2022). Specifically, we used an extreme gradient boosting (XGBoost)-based machine-learning method to predict the crime occurrence possibilities driven by crime and environmental variables, and the SHapley Additive exPlanations (SHAP) technique to interpret prediction results.

Furthermore, the location setting of crime prevention facilities, such as police stations and closed-circuit television (CCTV), is a primary crime prevention measure. This measure is crucial because the arrangement of these facilities can deter crime and reduce the fear of crime among residents. However, because of the cost and management involved, a limited number of these types of facilities can be developed. Despite the arrangements of crime prevention facilities requiring scientific research, limited studies have focused on the relationship between crime prevention facilities and spatial structures, such as the analysis of the appropriate locations of these facilities and their corresponding effectiveness. Studies have revealed that the appropriate arrangement of crime mitigation facilities can positively affect surrounding areas and crime prevention effects (Caplan et al. 2015; Troy, Grove, and O'Neil-Dunne 2012). Furthermore, the facilities tend to be arranged without consideration of factors influencing crimes. Residents typically demand the installation of these types of facilities in residential areas rather than commercial districts or station areas. Studies have speculated that although these facilities can provide psychological relief, they may not be effective in crime prevention (Jang, Kim, and Lee 2014; Lee and Kang 2012; Park and Choi 2009). Thus, identifying the relationship between crime occurrence, urban characteristics, and spatial structure to install crime prevention facilities in appropriate areas is critical.

Therefore, we applied machine-learning techniques to identify the relationship between the locations of crime reports and the surrounding independent variables (such as LULC) and examined the contributors to crime occurrence. First, we selected Daegu in the Republic of Korea as the study area and focused on urbanized areas. We then created a crime prediction map based on a model constructed using machine-learning techniques. We tested various machine-learning techniques (e.g., random forest, XBoost, CatBoost, and lightGBM) for a crime prediction model. Subsequently, we determined whether the corresponding hotspot areas have crime prevention facilities such as CCTV.

This study focuses on the following aspects. First, we examined which spatial selections of the independent variables (including LULC) can create a model with superior explanatory power. Second, we tested various machine-learning techniques to determine which machine-learning model yields the best fit for the corresponding data. Third, we identified which independent variables (including LULC) are highly likely to affect crime reports or 112 calls. Fourth, we confirmed whether the model constructed in this study can be used to create a crime risk map. Finally, we used the map to confirm whether crime prevention facilities (such as CCTV) are appropriately installed in areas with a high likelihood of crime occurrence.

This study identifies the intertwined influences between independent variables in urban environments, such as LULC and crime reports (112 calls). These provide valuable insights into the environmental conditions associated with crime occurrence. The results of these insights are expected to have a significant impact on urban planning and crime prevention policy decisions aimed at creating safer cities. For example, in existing cities, where the probability of crime occurrence is high, specific locations for installing crime prevention facilities such as CCTV or evidence collection systems can be proposed. In the case of new cities, the experimental results of this study could be used during the urban planning process to plan the installation of facilities like CCTV and police substations in areas where independent variables highly correlate with crime occurrence.

2. Materials and methods

2.1. Study area

This study focused on Daegu, the fourth highest populated city in the Republic of Korea, covering an area of 1499.5 km^2. Located in the Daegu Basin and surrounded by mountains, Daegu is a densely populated urbanized area and the third largest city in terms of urban population, following the Seoul Metropolitan and the Busan-Ulsan areas. As of December 2019, Daegu had a registered population of 2,468,222, among which 2,438,031 citizens were Korean. The city contains 1,031,251 households, with an average of 2.36 people per household (source: https://www.daegu.go.kr/english/index.do). In 2019, approximately 20 % of Daegu's area was urbanized, 10 % was agricultural, and approximately 53 % was forested (https://egis.me.go.kr/intro/land.do). For this study, a land cover map of Daegu, updated annually by the Korean Ministry of Environment, was used (Figure 1). We specifically extracted land cover data for urbanized areas designated for residential, industrial, commercial, cultural, sports and recreation, transportation, and public facilities to conduct our analysis.

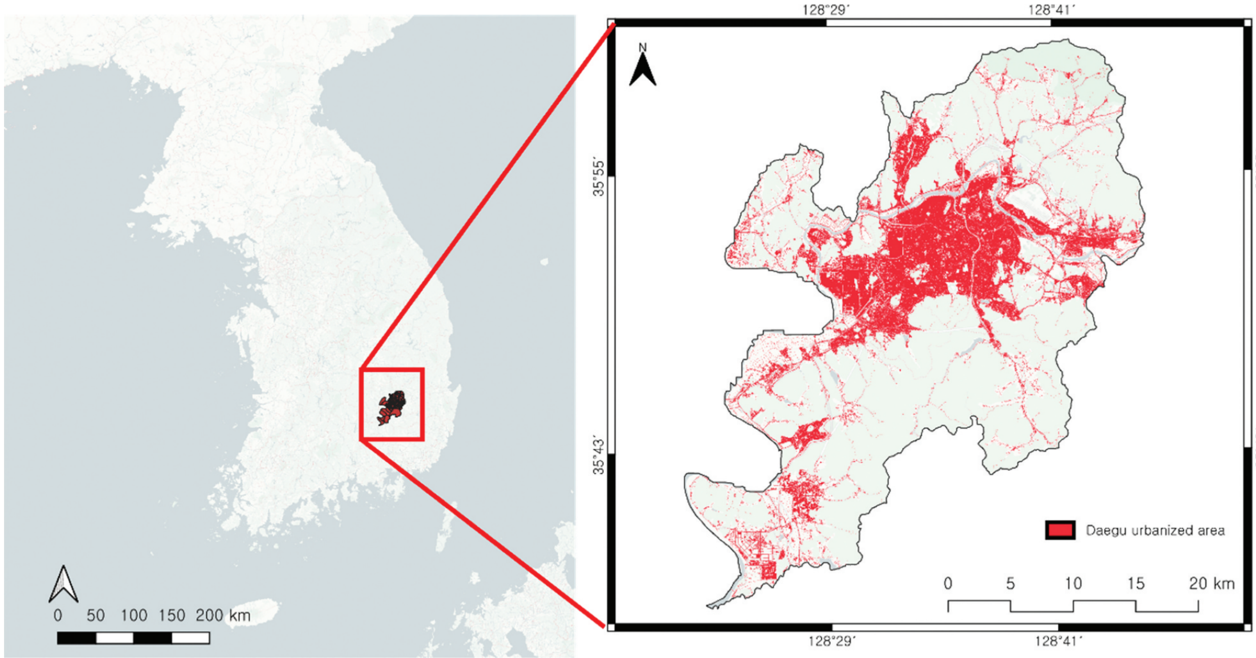


Figure 1. Study area: urbanized Daegu areas.

Table 1. Variables used in this study.

Data			Spatial Resolution	Source (Year)
Dependent variables	Crime location (112 calls)		Point data	Daegu Metropolitan City, Daegu Metropolitan Police Agency (2019)
Independent variables	Bus stop Closed-circuit television (CCTV) Police substation		Point data	Daegu Metropolitan City, Daegu Metropolitan Police Agency (2019)
	Point of interest	Library Elementary school Middle school High school University Adult entertainment establishment	Point data	National Geographic Information Institute (2019)
	Land cover area with detailed classifications	Urbanized area Agricultural area Forest area Grassland area Wetland area Bareland area Water area	1 m × 1 m	Land cover area with detailed classifications (2019)
	Statistical data	Number of households Number of houses Number of businesses Number of workers Population	100 m ×× 100 m	Statistical Geographic Information System (SGIS) (19.12.31)
	Floating population	Visiting Working House	50 m ×× 50 m	KT (19.07.31– ~20.02.28)
	Card sales		Polygon data	

2.2. Data

This study employed various independent variables to predict crime areas in Daegu. Table 1 summarizes the dependent and independent variables used. The dependent variable is the location of crimes reported in 112 emergency calls. The independent variables are:

Bus stops are frequent gathering places that may increase the likelihood of crimes. Similarly, CCTV and

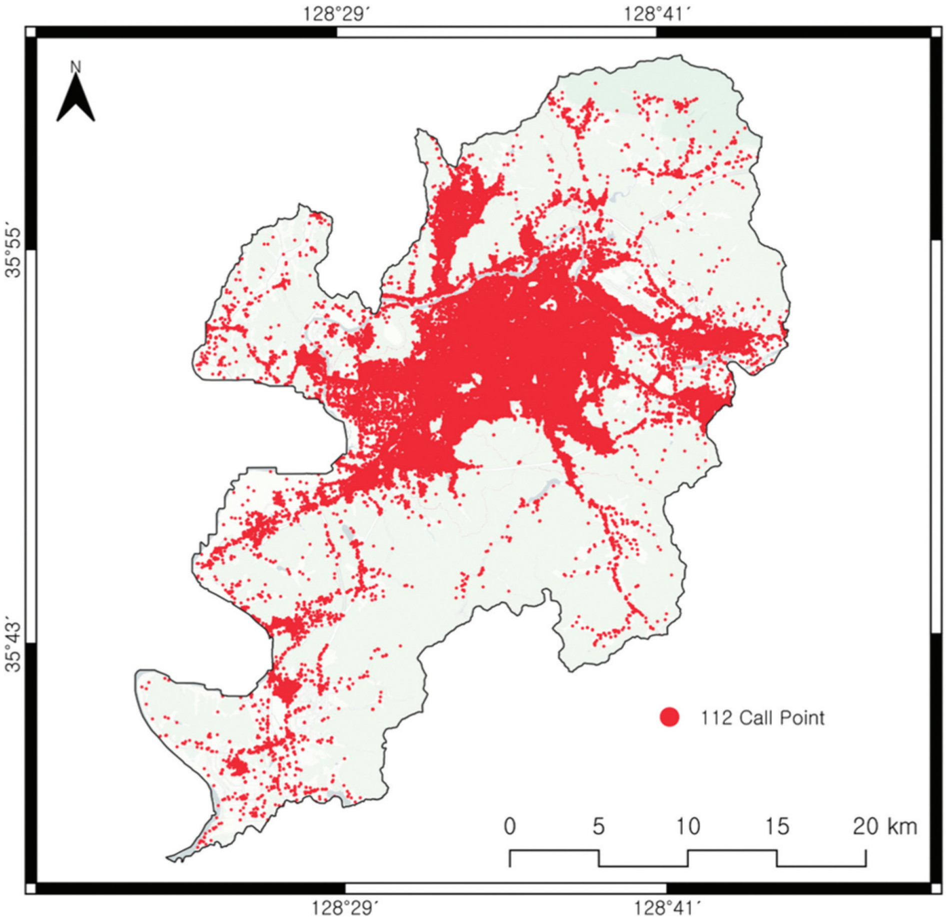


Figure 2. Maps of 112 call points in Daegu.

police substations act as deterrents to criminal activity and facilitate rapid responses, significantly impacting crime rates (Sherman, Gartin, and Buerger 1989; Skogan 2004; Tengbeh 2006). Points of interest (POIs), such as libraries, schools, and other locations associated with social and commercial activities, can correlate with higher crime rates in these areas (Lipton et al. 2013; Tabangin, Flores, and Emperador 2008). The detailed classifications of land covers reflect the geographic and environmental characteristics that influence crime patterns, with variations depending on land use types, such as commercial, residential, or transport hubs (Sypion-Dutkowska and Leitner 2017). Additionally, data on population density and economic activity levels, including statistical data, floating population, and card sales, provide crucial insights for assessing crime risk in specific areas (Boggs 1965; Schmid 1960a, 1960b; Witt, Clarke, and Fielding 1999).

2.2.1. Dependent variables

In this study, we utilized the 112 call location data provided by the Daegu Metropolitan City and the Daegu Metropolitan Police Agency, covering the period from July 2019 to 28 February 2020, to predict crime occurrence locations in urban areas (Figure 2).

In South Korea, 112 is the emergency number for the police. The 112 call data from Daegu include reports on various incidents, such as violent crimes, traffic offenses, minor infractions, and public order maintenance. To create a predictive model, this study incorporated all types of 112 call data.

One challenge associated with the use of machine-learning classification models to predict 112 call occurrences is data imbalance, which occurs when the number of instances in one class (e.g., non-112 calls) significantly exceeds those in another (e.g., 112 calls) (Batista, Prati, and Monard 2004). This imbalance can bias model training and impair its generalization capability. To mitigate this, we balanced the class ratio by generating approximately 370000 control points (non-112 calls), closely matching the 373387 recorded 112 call locations (Figure 3). In similar studies on 112 call data, employing a 1:1 sampling ratio has been shown to yield higher predictive accuracies (Corcoran 2019). Based on these findings, we adopted the same 1:1 sampling approach to ensure optimal predictive accuracy outcomes.

Each control point was positioned at least 10 m away from the nearest 112 call location and from other control points to prevent overlap and ensure a uniform spatial distribution. This balanced spatial sampling design enhances the representativeness of

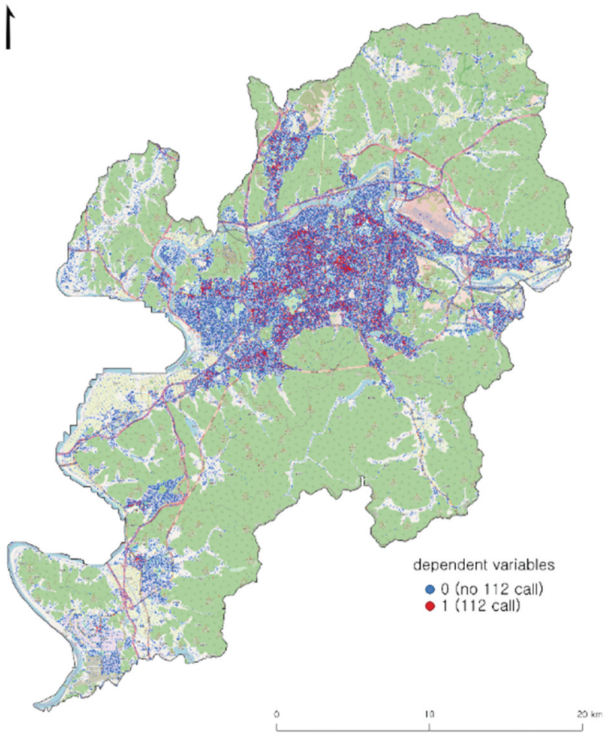


Figure 3. Dependent variables ((value: 1) = 112 calls, (value: 0) = control group).

the study area, improves statistical validity, and minimizes the risk of overfitting by ensuring the model learns without spatial bias (Theobald et al. 2007). By aligning the number of control points with the actual 112 call locations, we effectively addressed class imbalance in the model training process.

2.2.2. Independent variables

2.2.2.1. Bus stop, CCTV, and police substations.

The data for bus stops, CCTV locations, and police substations were obtained from the Daegu Metropolitan Police Agency for 2019. The bus stop dataset provides the geographic locations of bus stops in Daegu city. The CCTV dataset serves multiple purposes, such as crime prevention, child protection, traffic management, emergency monitoring, waste surveillance, and managing government buildings and cultural sites.

The data, initially in latitude and longitude coordinates, were converted into point SHP files and then transformed into grids with a resolution of 10 m (Figure 4).

2.2.2.2. POI. The POI data include libraries, elementary schools, middle schools, high schools, universities, and adult entertainment establishments. This dataset from the National Geographic Information Institute (NGII) was developed by integrating public administration databases for location-based services. The data were provided in point format (https://map.ngii.go.kr/mn/mainPage.do).

The NGII's POI data covers various categories, such as education (elementary, middle, and high schools, universities), culture (libraries), security (police substations), and others (adult entertainment establishments). For this study, the 2019 POI data from NGII were used. Initially provided as point SHP files (Figure 5), the data were converted into grids with a resolution of 10 m.

2.2.2.3. Land cover and land use with detailed classifications. The detailed land cover map of South Korea, a national statistical product, was provided by the Statistical Environmental Spatial Service. This map is a thematic representation at a 1:5000 scale, created by classifying land into 41 categories using imagery at a resolution of 1 m (https://egis.me.go.kr/intro/land.do). For the Daegu region, 35 of these categories were utilized based on 2019 data, excluding those not present in the area, such as Airport (151), Port (152), Mudflat (521), Salt Farms (522), Beach (611), and Marine Water (721). For this study, the land cover data were converted to a resolution of 10 m × 10 m (Figure 6 and Table 2).

2.2.2.4. Statistical data. The statistical data were obtained from the 2019 Statistical Geographic Information System (SGIS), managed by Statistics Korea. SGIS is an information system designed to aid decision-making by providing data that combines population, housing, and business census information with geographic attributes, such as coordinates, boundaries, and maps (https://sgis.kostat.go.kr/view/index). SGIS offers aggregated statistical data for each 100-meter grid, including totals for households, house units, population, businesses, and workers.

In SGIS terms, a "household" is defined as a unit where individuals live together, sharing cooking and sleeping arrangements. A "house" is a building used for living, with at least one room, a kitchen, and an independent exit. "Population" is the number of people in a specific area. "Business" refers to any enterprise that conducts economic activities independently in a particular location. "Worker" denotes individuals engaged in specific types of work (https://sgis.kostat.go.kr/mobile/board/term.sgis).

For this study, we used the 2019 grid statistical data from SGIS. The data, initially in 100-m grid SHP files, were converted into raster files (at a resolution of 10 m) using an area-weighted average method (Figure 7).

2.2.2.5. Floating population. The KT big data platform provides floating population data to analyze the actual living population, rather than traditional

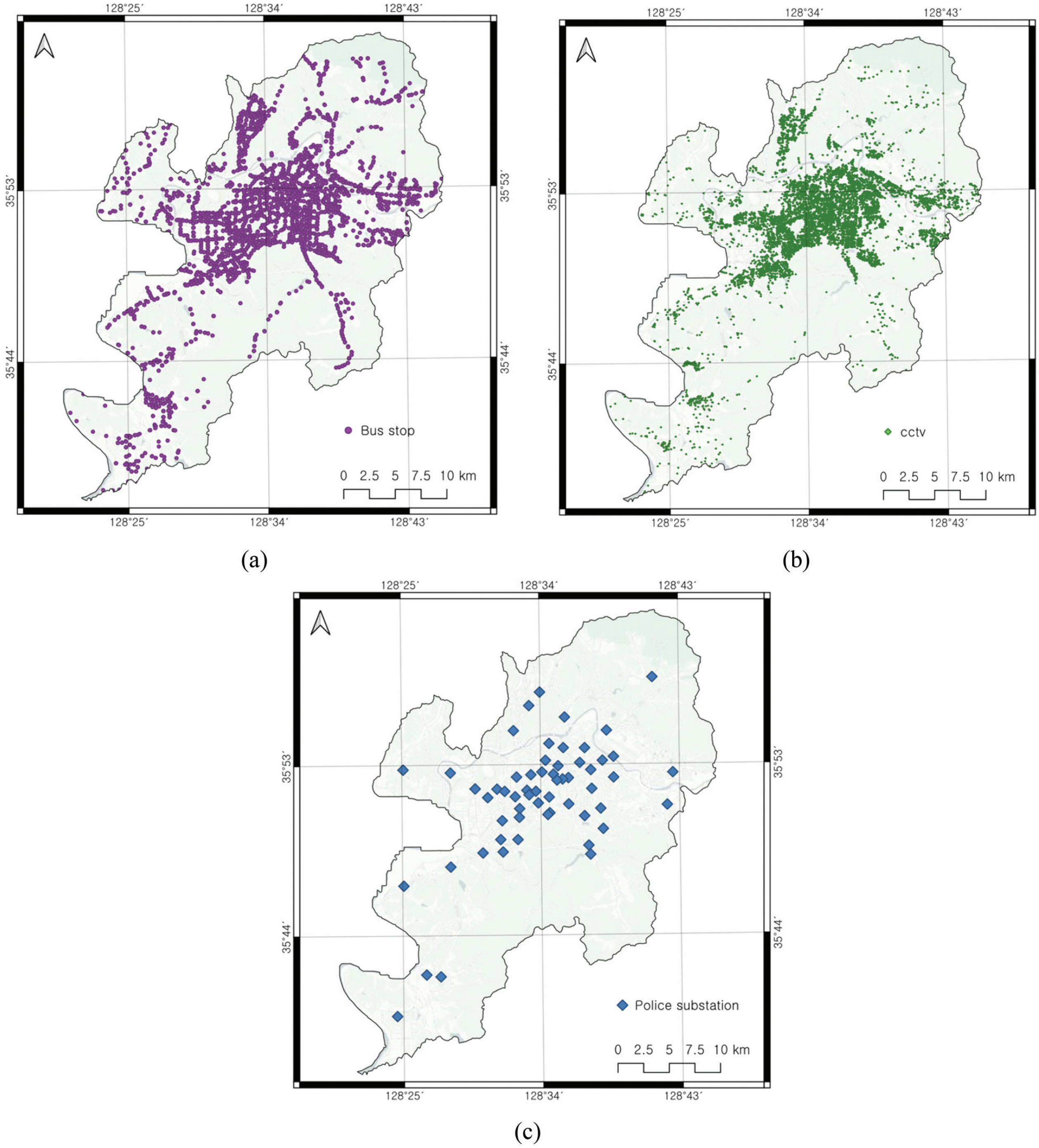


Figure 4. Location point data of (a) bus stops, (b) closed-circuit televisions (CCTVs), and (c) police substations.

address-based demographics (https://enterprise.kt.com/pd/P_PD_AI_BD_003.do). KT's floating population data reflects the total movement of individuals, excluding residents, within a specific area, allowing for multiple counts within an hour.

This study focused on the domestic population – specifically visitors, workers, and house – to examine their relationship with 112 emergency calls. The data, collected from July 2019 to February 2020, were converted from average daily floating population figures into point data and then aggregated into raster files (at a resolution of 10 m) for total sum calculations (Figure 8).

2.2.2.6. Card sales. The KT big data platform also provides card sales data. Available from 31 July 2019, to 28 February 2020, this data is in CSV format, containing hourly card sales records for various businesses, each linked to its administrative district. To process the data, we first calculated the total daily card sales for each business and then aggregated these totals at the district level. The aggregated sales data were assigned to the corresponding district shapefile and divided into 10 m × 10 m grids. The value for each grid cell was calculated as the ratio of total card sales in the district to the district's area, allowing for the use of area-weighted averages in zonal statistics (Figure 9).

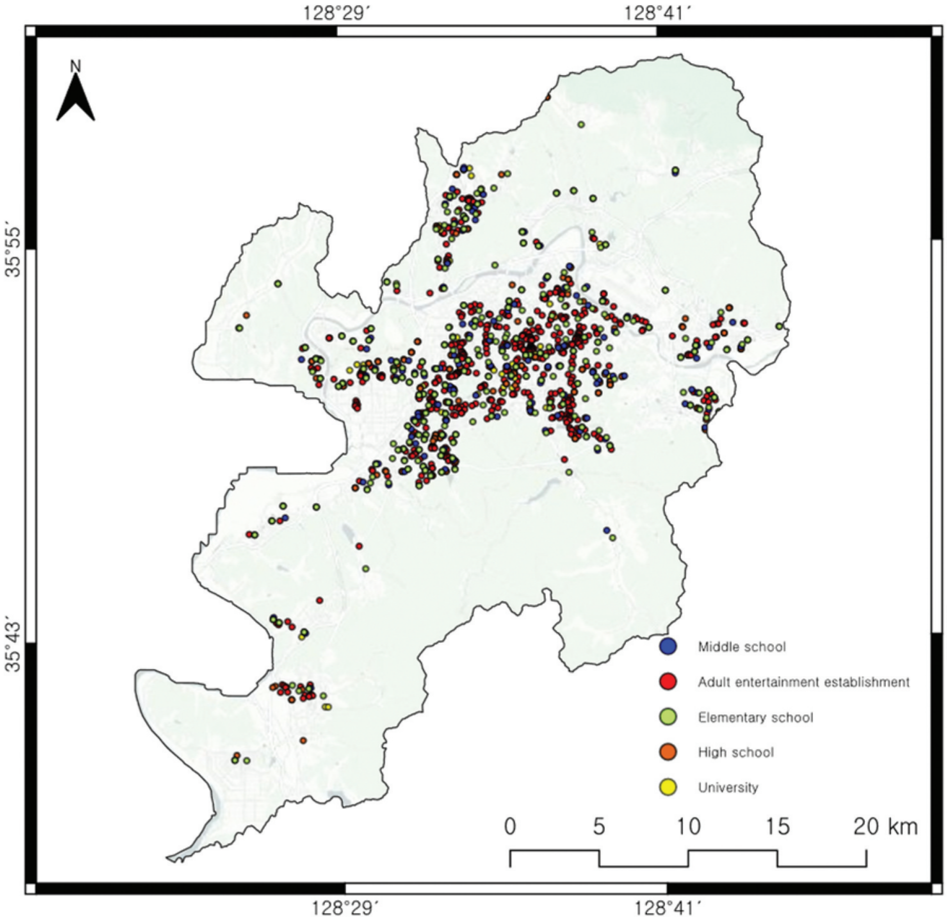


Figure 5. Point-of-interest (POI) data.

2.3. Methods

2.3.1. Research procedure

The study utilized various datasets, including 112 call records, land cover, bus stops, CCTV, police substations, POIs, statistical data, floating population data, and card sales. Zonal statistics were applied to create structured data for a classification model. To identify the optimal model, we compared the accuracy and F1 scores of four tree-based machine-learning models using datasets with buffer distances ranging from 10–50 m. Hyperparameter tuning was conducted via a random search to optimize the model parameters. Additionally, SHAP value analysis from XAI was employed to assess the importance and relationships of each independent variable. The final model was then used to predict high-risk crime areas lacking nearby CCTV coverage in Daegu (Figure 10).

2.3.2. Comparison of machine-learning models

We used random forest and boosting models. Specifically, the models XGBoost, CatBoost, Random Forest, and LightGBM were used for analyses. Figure 11 shows the results and compares the accuracy and F1 scores of all models for different buffer distances. XGBoost yields the highest accuracy and F1 scores that peak at 20 m. Therefore, these models performed best at 20 m. As a result, we used 20-m zonal statistics to construct the structured data and applied XGBoost for further analysis.

2.3.3. XGBoost model

Following our model comparison, XGBoost was selected for further analysis owing to its superior performance at a 20-m buffer distance. XGBoost, proposed by Chen and Guestrin (2016), is an advanced gradient-boosting model designed for speed, scalability, and accuracy (Chen and Guestrin 2016). It functions as a boosting classifier that combines multiple weak tree models to create a strong predictive model with high accuracy and low-error rates through iterative improvements.

The model's prediction for each instance ($\widehat{y}_i$) is expressed as,

$$\widehat{y}_i = \sum_{k=1}^{K} f_k(X_i), f_k \in F, \tag{1}$$

where f_k represents each tree in the model, X_i is the input variable, and F is the function space of all classification trees (CART).

XGBoost optimizes the following objective function,

$$Obj = \sum_{i=1}^{n} l(y_i, \widehat{y}_i) +, \sum_{k=1}^{K} \Omega(f_k), f_k \in F. \tag{2}$$

This objective function comprises two components: the training loss function l, which measures the

Table 2. Detailed classifications (35 subdivisions).

Classification Item Name	Classification Code
Single-detached residential facility	111
Apartment housing facility	112
Industrial facility	121
Commercial/Business facilities	131
Areas with mixed purposes	132
Culture/Sports/Recreational facilities	141
Railroad	153
Road	154
Other transportation modes/Communication facilities	155
Basic environmental facilities	161
Education and administrative facilities	162
Other public facilities	163
Rice field with redevelopment of arable land	211
Rice field without redevelopment of arable land	212
Farmland with redevelopment of arable land	221
Farmland without redevelopment of arable land	222
Protected cultivation	231
Orchard	241
Ranch/Aquafarms	251
Other cultivation areas	252
Broadleaf forests	311
Coniferous forests	321
Mixed forests	331
Natural grassland	411
Golf course	421
Cemetery	422
Other grasslands	423
Inland wetlands (Waterfront vegetation)	511
Riverbanks	612
Rocks	613
Mining site	621
Playground	622
Other bare lands	623
River	711
Lake	712

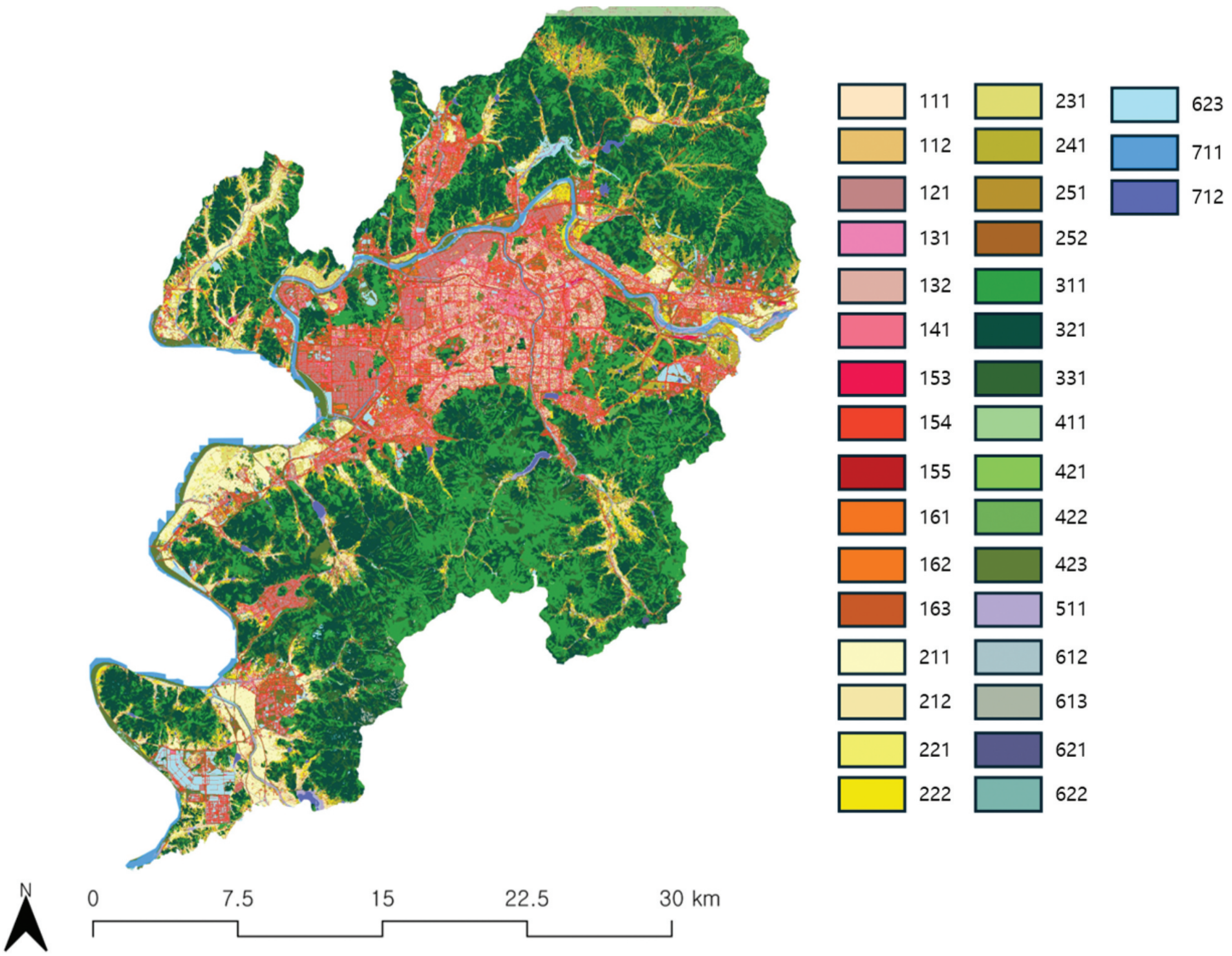


Figure 6. Land cover map of Daegu with detailed classifications.

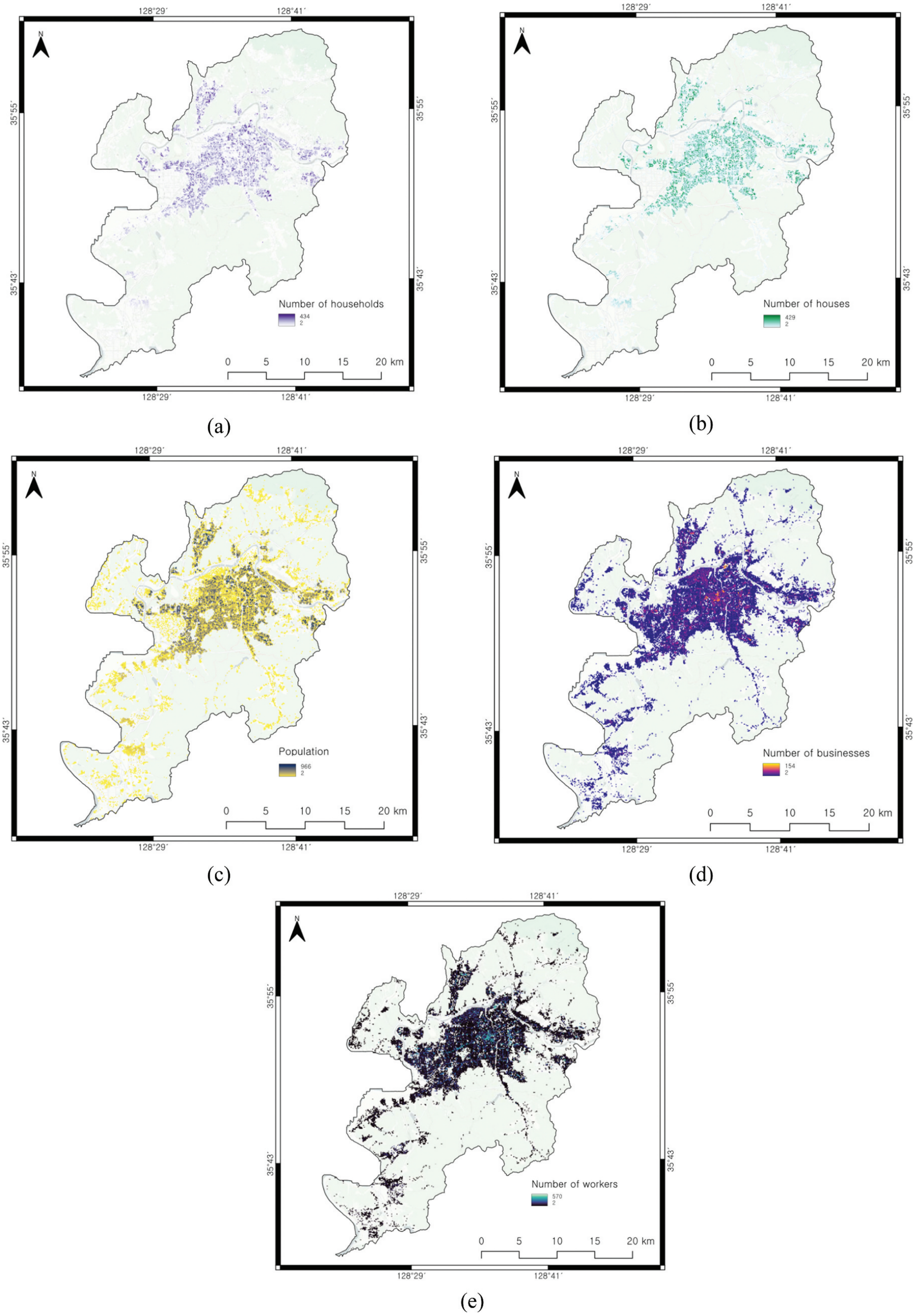


Figure 7. Statistical data: (a) Households; (b) Houses; (c) Population; (d) Businesses; and (e) employees.

difference between the predicted ($\hat{y}_i$) and actual (y_i) values, and the regularization term Ω, which prevents overfitting by assessing model complexity (Meng et al. 2021).

Multicollinearity refers to the regression model cases wherein one predictor variable can be linearly predicted from others, thus leading to biased or misleading results (Farrar and Glauber 1967). To

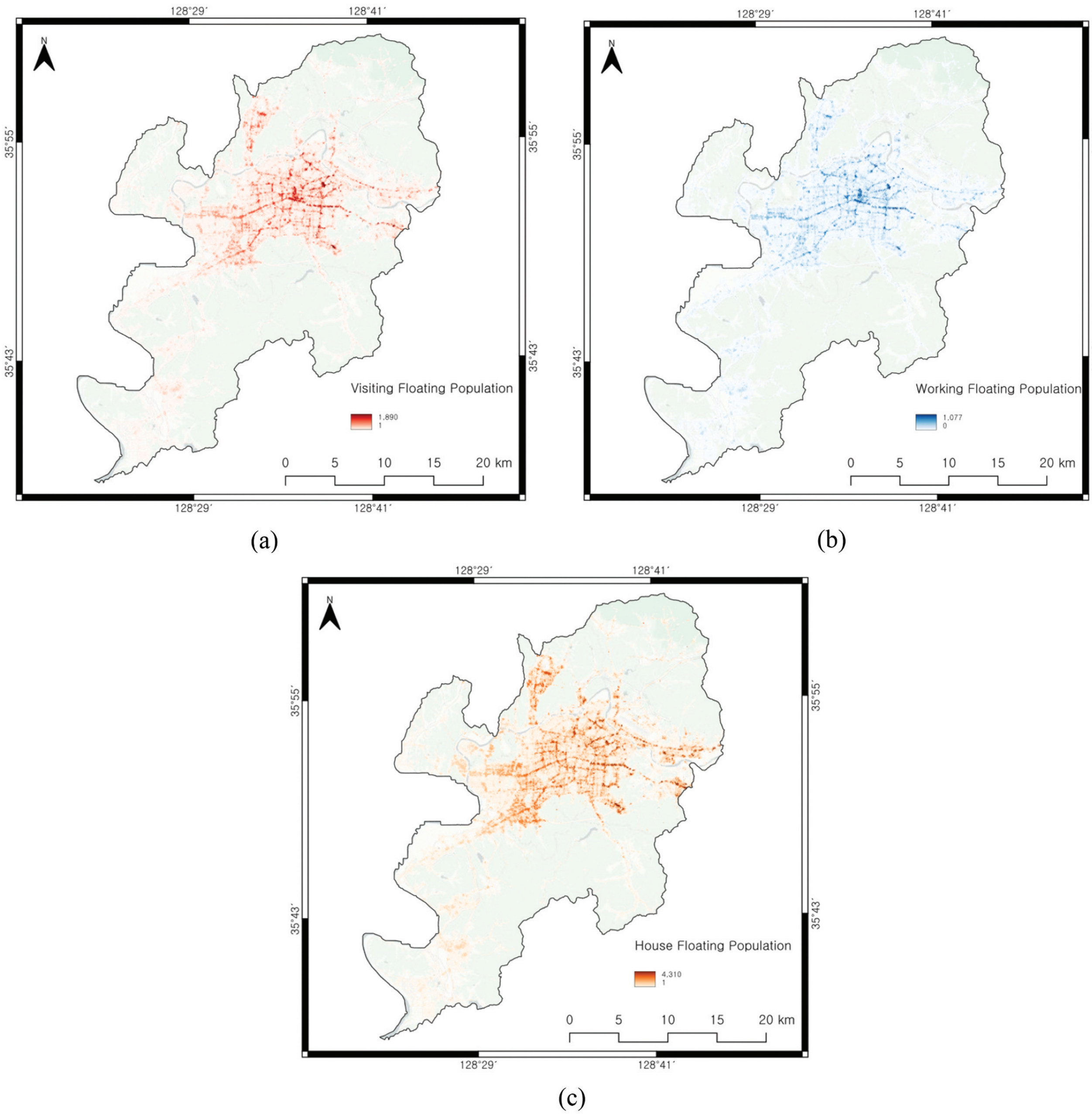


Figure 8. Floating populations: (a) visiting population; (b) working population; and (c) house population.

address the issue of multicollinearity, regularization techniques can be employed. In this study, these techniques were applied to the machine-learning models. Additionally, decision tree and boosting tree algorithms, particularly the XGBoost model, effectively handle multicollinearity by selecting the optimal predictor variables during the modeling process (Chen and Guestrin 2016).

2.3.4. SHAP

SHAP is a model interpretation technique that helps explain complex machine-learning models, such as ensemble methods and DNNs. It provides approximate values that indicate how each feature impacts a model's predictions in an interpretable way (Slack et al. 2020). SHAP is model-agnostic; this means that it can be used with any machine-learning model to interpret individual predictions.

SHAP works by evaluating the effects of all the independent variables on the predicted outcome. In a dataset D with N data points $(X, y) = \{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\}$, each x_i is a vector of feature values, and y_i is the corresponding class label. The model f is a black-box classifier that outputs a class label for a given data point. SHAP uses an explanatory model g (which is a linear approximation of f) to understand how the features contribute to the predictions. The complexity of the model g is quantified by $\Omega(g)$, the number of nonzero weights in the linear model. SHAP enables the interpretation of the effects of features based on game theory-based principles.

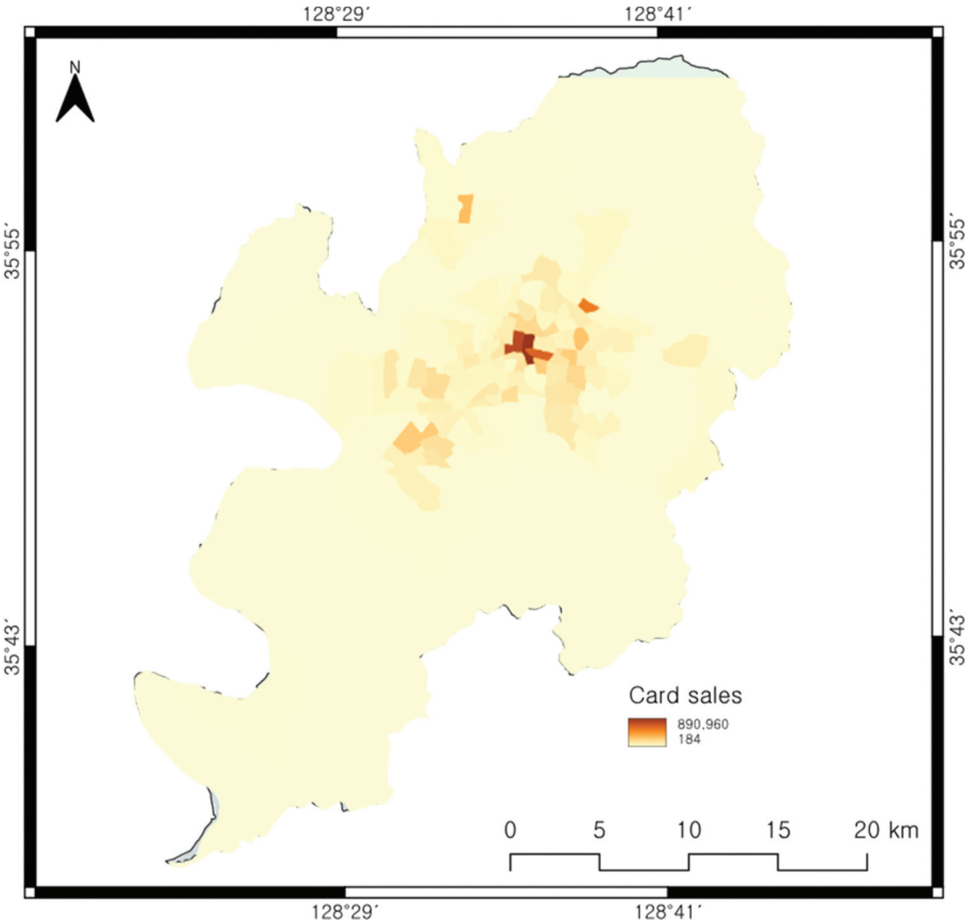


Figure 9. Card sales.

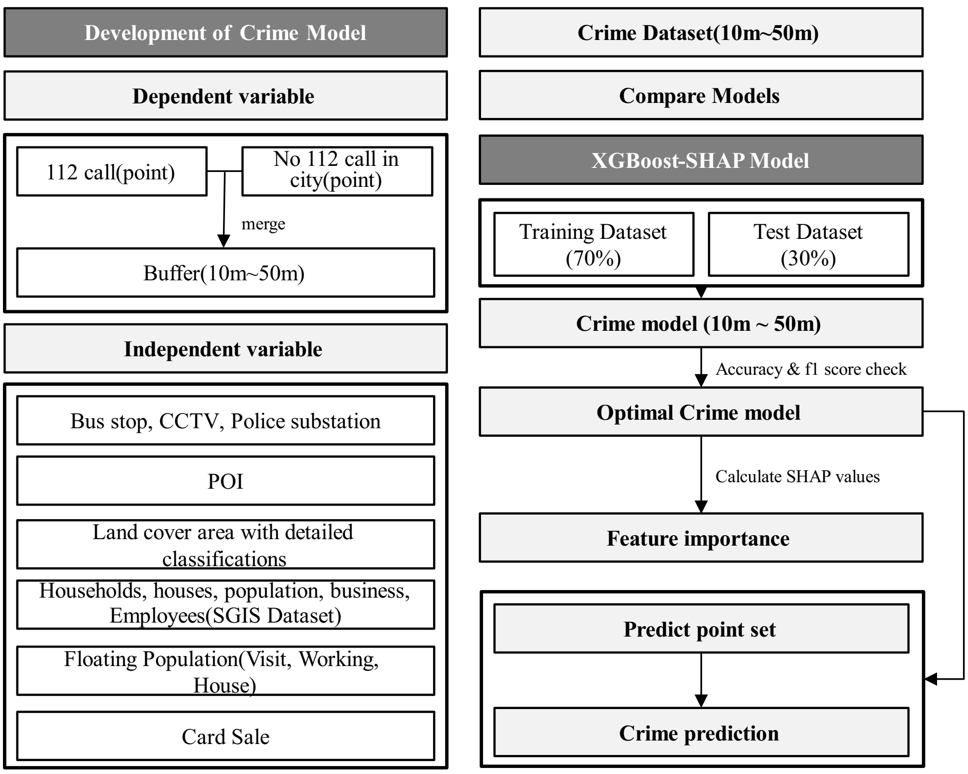
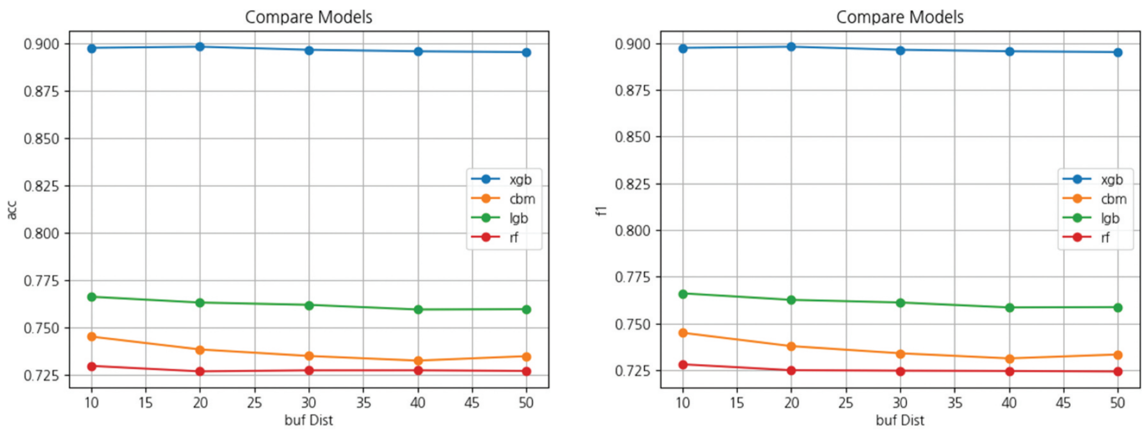


Figure 10. Flow chart summarizing the research process adopted in this study.



(a) (b)

Figure 11. Comparison of tree-based models (XGBoost, Catboost, LightGBM, and random forest): (a) accuracy and (b) F1 score.

Table 3. Hyperparameter ranges used for tuning.

Hyperparameter ranges used for tuning	XGBoost	*'''subsample':'' uniform(0., 0.3), '''reg_lambda':'' uniform(0, 0.1), '''reg_alpha:''': uniform(0, 0.1), '''n_estimators:''': randint(1000, 3000), '''max_depth:''': randint(5, 10), '''learning_rate:''' : uniform(0.05, 0.15), '''gamma:''': uniform(0, 1), '''colsample_bytree:''' : uniform(0.4, 0.6),*

Table 4. Hyperparameter tuning results.

Parameters	Values
Subsample	0.9049763486783569
Reg_lambda	0.06775643618422825
Reg_alpha	0.05908929431882418
N_estimators	2794
Max_depth	9
Learning_rate	0.12622981651110782
Gamma	0.15071754396542947
Colsample_bytree	0.8790070749907306

3. Results

3.1. Hyperparameter tuning

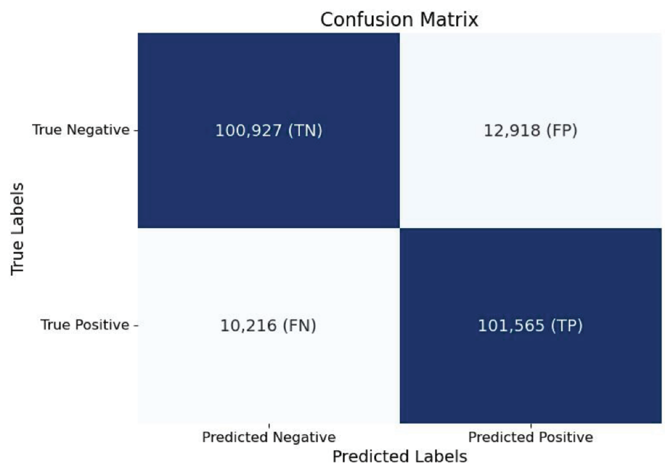
We used XGBoost with a 20-m buffer based on the statistical values of the independent variables. To improve model accuracy, we applied hyperparameter tuning using the random search method. Table 3 shows the range of hyperparameters tested, and Table 4 presents the final hyperparameter settings. The model achieved an accuracy of 0.897.

To evaluate the model's performance with the tuned hyperparameters, we calculated four metrics: accuracy, precision, recall, and F1 scores (Figure 12(a)). These metrics were derived from the confusion matrix (Figure 12(b)).

The final model for predicting 112 calls had an accuracy of 0.897, a precision of 0.898, recall of 0.897, and an F1 score of 0.897. According to the confusion matrix, the model correctly predicted 101,565 true

Indicators	Values
Accuracy	0.897
	0.898
Recall	0.897
F1 Score	0.897

(a)



(b)

Figure 12. (a): accuracy, precision, recall, and F1 score outcomes; (b) confusion matrix.

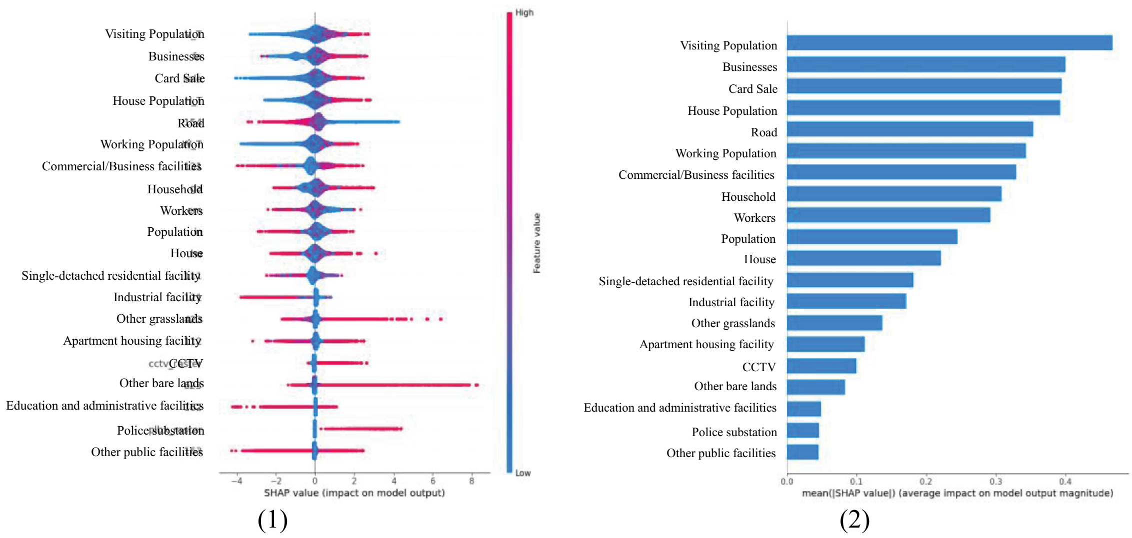


Figure 13. Shapley additive exPlanations (SHAP) values. (1) direction of SHAP values: SHAP values to the right indicate a positive influence on the likelihood of 112 calls, while values to the left indicate a negative influence. The color gradient (red to blue) represents the magnitude of each feature's value. (2) relative importance: the average SHAP value magnitude for each feature, indicating its overall contribution to the model predictions.

positives and 100,927 true negatives, with 12,918 false positives and 10,216 false negatives (Figure 12).

3.2. SHAP analysis

We used the XGBoost model with 20-m zonal statistics to train the structured data and conducted a SHAP value analysis to predict crime occurrences. SHAP analysis helps evaluate how independent variables impact model predictions, identifying the importance and direction of influences of these variables in complex models (Figure 13).

The SHAP analysis revealed that the most important factors for predicting crime reports were the visiting population, businesses, card sales, house population, roads, and working population (in this order). The visiting population, businesses, house population, and working population were positively associated with crime reports, suggesting that higher numbers in these categories increase the likelihood of crime. Card sales, which indicate the economic activity, also showed a positive relationship with crime reports, implying higher crime rates in areas with higher card usage. Conversely, roads exhibited a negative relationship with crime reports, suggesting that well-maintained roads may reduce crime by enhancing public safety.

3.3. Crime prediction

The 112 call prediction indicates that the probability of 112 calls is higher in the city center. Specifically, the probability of 112 call prediction is similar to the visiting population, as can be inferred from the SHAP values (Figures 14 and 15).

Based on the comparison of the 112-call prediction map and CCTV locations, Figure 16 shows that Daegu was mostly well equipped with CCTVs in areas with a high probability of 112 calls. However, as presented in Figure 16 (blue box), areas to be considered could be derived for the installation of additional facilities such as CCTVs.

4. Discussion

Various machine-learning techniques and artificial neural network models have been previously employed to analyze the relationship between urban features, urban spatial structure, and crime occurrence and predict crime occurrence locations. These techniques are particularly useful for identifying areas with a high likelihood of crime occurrence. However, these methods are insufficient for interpreting and explaining why the specific areas record higher crime rates, which is the limitation of using black box models. Therefore, we selected a model with excellent explanatory power among representative machine-learning techniques to overcome the shortcomings of black-box models. We also used SHAP, which is a representative model among XAI techniques, to analyze the relationship between crime occurrence, land use in the surrounding area, locations of main facilities, as well as urban statistics (e.g., population, houses, and businesses), floating population, and card sales data. The analyzed results were then used to predict the areas with high likelihoods of crime occurrence.

Comparing representative machine-learning techniques based on random forest and boosting series techniques revealed that the XGBoost model was the

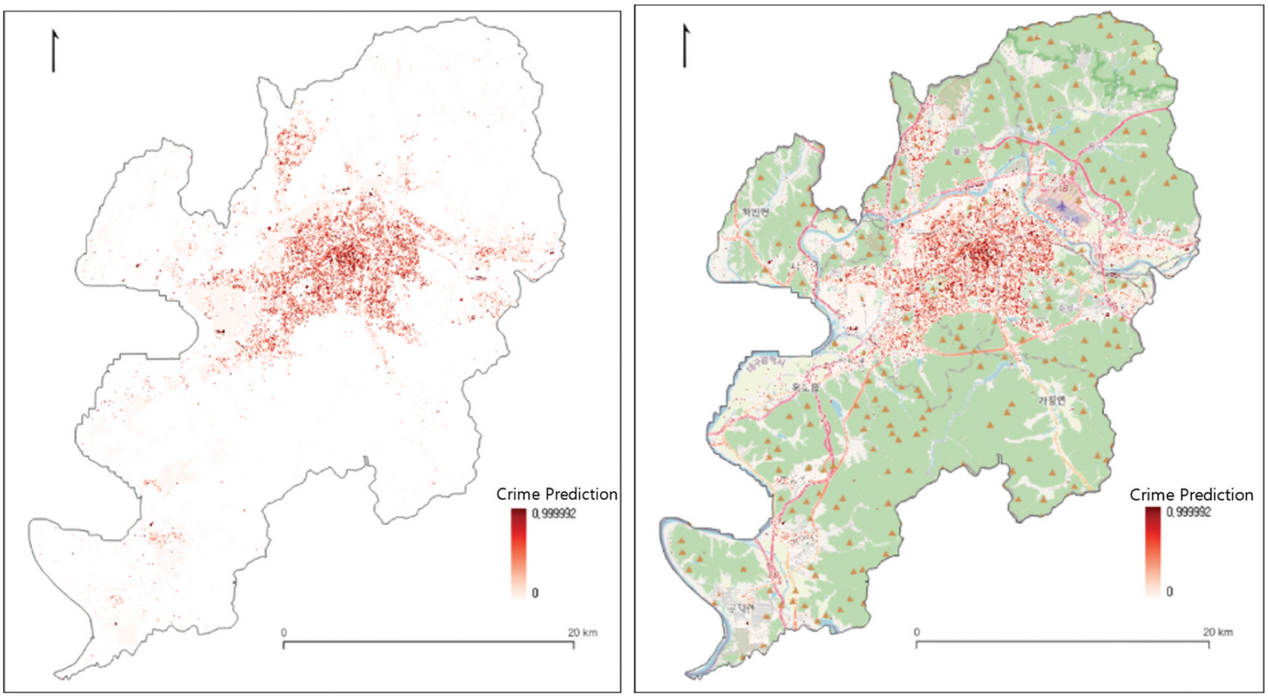


Figure 14. Prediction maps.

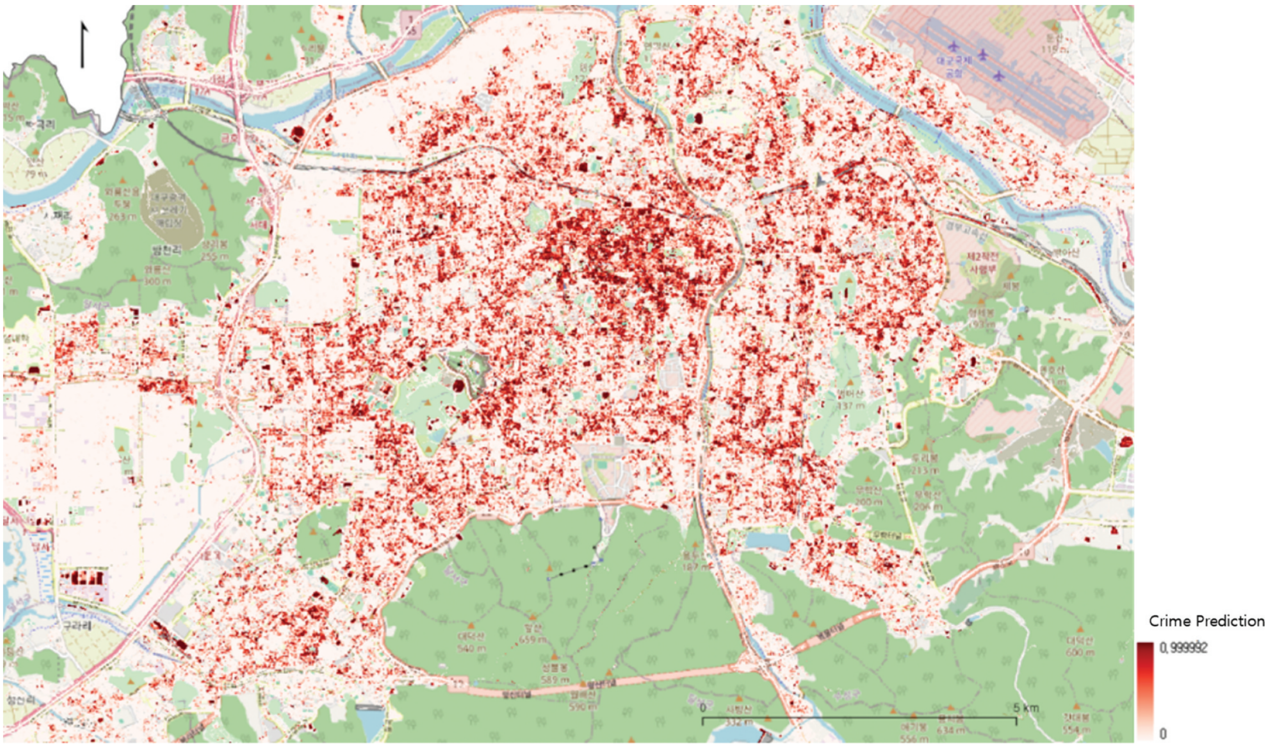


Figure 15. Enlarged version of a map showing 112 call predictions.

most efficient; the explanatory power was excellent when using independent variables included in the corresponding space based on a 20-m buffer around the center of each raster (considering eight 20-m grids in the surrounding area). SHAP analysis indicated that the visiting population, roads, commercial districts, card sales, and the number of households (in this order) highly affected crime reports (112 calls). Based on this model, we created a crime risk map or a probability map of crime occurrence for Daegu Metropolitan City. Without using CCTVs, this probability map revealed areas with higher probability of crime occurrence than other areas.

The analyzed results revealed that the XGBoost model was the optimal model in this analysis compared with other boosting models, namely LightGBM, CatBoost, and random forest models. The results indicate that XGBoost incorporates the advantages of

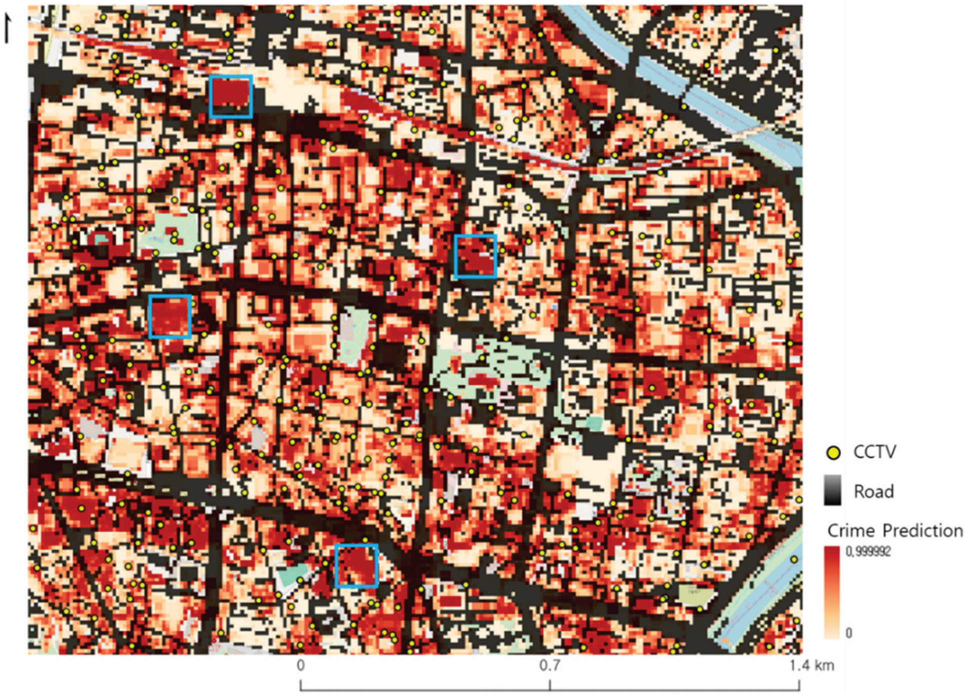


Figure 16. 112 call prediction map and CCTV status.

gradient boosting and overcomes problems such as overfitting and learning speed. These results are similar to the findings of other studies that have confirmed its good performance and excellence in classification, ranking learning, and prediction tasks. Although studies that conducted a comparative analysis of XGBoost, LightGBM, CatBoost, and random forest models (Alves, Ribeiro, and Rodrigues 2018; Kim and Lee 2023; Tong et al. 2021; Wheeler and Steenbeek 2020) have indicated that XGBoost is the optimal model, other models have also been selected as optimal models but for other applications (Kim and Lee 2023). Therefore, considering the data characteristics of the dependent and independent variables is crucial when selecting optimal models.

Second, the highest explanatory power was obtained when a buffer space of 20 m was considered in the surrounding area based on the analysis of the spatial range, including various factors that affected 112 calls. The explanatory power of the machine-learning models applied in this study slightly decreased as the buffer space used for analysis increased. These outcomes indicate that the explanatory power of each machine-learning model differs considerably, but the accuracy trend according to the statistical values of the independent variables included in each considered space revealed that there is a space to be considered according to the analysis. Therefore, because limited studies have investigated the explanatory power of the model according to the distance from the location of the crime occurrence, the findings of this study can be considered significant for urban planning factors to prevent crime.

Furthermore, visiting population, businesses, card sales, house population, road, working population, and the areas of commercial/business facilities, which were included within 10 m from the occurrence points, considerably affected the 112 calls. Moreover, areas (especially commercial districts) that have high visiting, working, and house population, areas with increased numbers of businesses (i.e., shops) and card sales, and locations away from roads are highly likely to be associated with increased crime rates. Therefore, installing facilities to prevent crime in such areas is critical. Such analytical outcomes are consistent with those of previous studies that have revealed a high relationship between commercial centers and crime occurrences (He and Li 2021; Twinam 2017; Weisburd et al. 2020). We obtained the same results as those obtained in a prior study that showed that as the residential areas increase, the crime occurrence also increases (Sampson 1983). In particular, our results correlate with those of a study that showed a positive relation between the mixed uses of residential and commercial spaces and crime occurrence (Yirmibesoglu and Ergun 2007). Furthermore, previous studies could not effectively analyze the spatial independent variables that influenced the dependent variables by making predictions related to the 112 calls. This study overcame this shortcoming and yielded results by considering the spatial factors that influenced the 112 call predictions.

Fourth, we constructed a 112-call prediction map of Daegu Metropolitan City's urban areas by using the model built in this study. Similar to the SHAP analysis, the 112-call prediction map revealed high probabilities of crime occurrence around areas with high levels of visiting population, businesses, card sales, house population, working population, and areas of commercial/business facilities. Because of overlapping locations of existing CCTVs, CCTVs were observed in areas with high probabilities of the 112 call occurrence. However, the SHAP analysis did not reveal any significance in the absence or presence of CCTVs in the surrounding areas to affect the model's results. However, considering that CCTVs can prevent crime and collect evidence to solve incidents, the absence or presence of CCTVs is crucial in most areas where 112 calls are predicted to occur. Thus, installing new CCTVs at points where the probability of 112 calls is high, as suggested in this study.

5. Conclusions

This study holds significant value in comprehensively analyzing the relationship between various urban environmental factors and crime occurrences in Daegu using machine-learning techniques and in building a crime prediction model based on the findings. Using the XGBoost model adopted in this study, we identified the influences of variables such as visiting population, card sales, roads, and commercial district areas on crime occurrence. Additionally, based on SHAP analysis, we were able to explain intuitively the importance and impact of each variable. Notably, this study found that visiting population and card sales are major factors responsible for the increase in the probability of crime occurrences. Based on the predictions for high-crime areas, the need for new CCTV installations was suggested.

The analyzed results indicated that the probability of 112 calls was higher in areas with a large visiting population, card sales, house population, and working population. For example, the probability of 112 calls occurring in areas with a high visiting population was approximately 25 % higher, and in areas with active card sales, the probability was > 15 % higher. This aligns with previous studies (He and Li 2021; Sampson 1983) that suggested higher crime rates in commercial areas with increased foot traffic and active economic activities. By contrast, roads yielded a negative correlation with crime occurrence, with areas that had well-maintained roads that exhibited approximately 10 % lower crime occurrence probability. This finding is similar to those reported in previous studies (Lee and Kang 2012; Wolfe and Mennis 2012), suggesting that well-developed public infrastructure (such as roads) has a positive impact on law enforcement activities.

Based on the research model, a crime occurrence prediction map was created for the Daegu city, and SHAP analysis was used to evaluate the relative importance of crime occurrence factors, enabling a clearer explanation of crime risks in specific areas. The analysis of the prediction map revealed that while most high-crime probability areas in Daegu were equipped with CCTV, a small portion of areas (approximately 3 %) with a high probability of 112 calls did not have CCTVs installed. It is suggested that new CCTV installations should be made in these areas in the future, providing important implications for crime prevention and law enforcement activities.

This study contributes to the existing body of research that uses machine-learning techniques to analyze the relationship between urban environmental variables and crime occurrence in three prominent ways. First, through a comparison of machine-learning models, the superiority of XGBoost was demonstrated; this is expected to be helpful in future studies when reviewing related models. Second, more structured data were utilized, and a meaningful influence distance was examined by comparing buffer distances ranging from 10 m to 50 m, providing a more advanced perspective compared with those of previous studies. This aspect is also expected to contribute to future studies by offering a good starting point for research. Finally, the use of XAI techniques allowed for the interpretation of crime occurrence prediction results, providing policymakers with useful data when making decisions related to crime prevention. This approach is considered to overcome the limitations of traditional black-box models used in previous crime prediction research (Kim, McCarty, and Jeong 2023; Wang et al. 2020).

Despite its contributions, this study has certain limitations. One limitation is that it did not consider different types of crimes but instead used the overall 112 call data. Additionally, the study did not specifically analyze the time of the 112 calls, which could also be considered as a study limitation and a potential area for new research discoveries based on additional analyses.

Given these limitations, future research should aim to develop more refined crime prediction models by segmenting the analysis based on crime types. For instance, as property and violent crimes may be influenced differently by urban environmental factors, it would be important to analyze the occurrence patterns of each crime type in detail. Furthermore, applying the model used in this study to other cities or countries and conducting comparative analyses could enhance the external validity of the research. These additional studies could provide more detailed preventive strategies for urban planning and policy-making, ultimately contributing to the creation of safer and more sustainable urban environments.

Acknowledgements

The authors would like to thank the city of Daegu.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was funded by grants [21NSPS-B149944-04] from the Architecture and Urban Development Research Program funded by the Ministry of Land, Infrastructure, and Transport of the Korean Government.

Notes on contributors

Geunhan Kim is a specialist researcher in Korea Environment Institute. His research focuses on GeoAI and environmental planning.

Youngtae Cho is a research fellow at the Land and Housing Research Institute and the head of the Urban Research Group. Dr. Cho participates as a member of the National Smart Cities Committee and is also an adjunct professor at Korea University. His areas of expertise are urban development and smart cities.

Ju-Heon Lee is a student in Department of Geoinformatics at the University of Seoul. He participated in project research part-time at KEI and completed an internship at the GeoDatabase Lab.

Gunwon Lee is a professor in the Department of Architecture at Korea University, where he also serves as the chair of both the Program of Urban Regeneration and the Department of Smart City. His research focuses on urban design and climate/environmental planning.

References

Alves, L. G., H. V. Ribeiro, and F. A. Rodrigues. 2018. "Crime Prediction Through Urban Metrics and Statistical Learning." *Physica A: Statistical Mechanics and Its Applications* 505:435–443. https://doi.org/10.1016/j.physa.2018.03.084.

Andresen, M. A., and G. W. Jenion. 2008. "Crime Prevention and the Science of Where People are." *Criminal Justice Policy Review* 19 (2): 164–180. https://doi.org/10.1177/0887403407311591.

Bagwell, R., W. E. Leal, S. Sen Roy, H. Flanagan, L. Britton, A. R. Piquero, and K. Block. 2024. "The Geospatial Patterning of Crimes Against Persons Calls for Service on Days with and without San Antonio Spurs Games." *Journal of Experimental Criminology*. https://doi.org/10.1007/s11292-023-09605-6.

Batista, G. E., R. C. Prati, and M. C. Monard. 2004. "A Study of the Behavior of Several Methods for Balancing Machine Learning Training Data." *ACM SIGKDD Explorations Newsletter* 6 (1): 20–29. https://doi.org/10.1145/1007730.1007735.

Boggs, S. L. 1965. "Urban Crime Patterns." *American Sociological Review* 30 (6): 899–908. https://doi.org/10.2307/2090968.

Brantingham, P. J., and P. L. Brantingham. 1981. *Environmental Criminology*. Beverly Hills, CA, USA: Sage Publications.

Briz-Redón, Á., J. Mateu, and F. Montes. 2021. "Identifying Crime Generators and Spatially Overlapping High-Risk Areas Through a Nonlinear Model: A Comparison Between Three Cities of the Valencian Region (Spain)." *Statistica Neerlandica* 76 (1): 97–120. https://doi.org/10.1111/stan.12254.

Browning, C. R., R. A. Byron, C. A. Calder, L. J. Krivo, M.-P. Kwan, J.-Y. Lee, and R. D. Peterson. 2010. "Commercial Density, Residential Concentration, and Crime: Land Use Patterns and Violence in Neighborhood Context." *The Journal of Research in Crime and Delinquency* 47 (3): 329–357. https://doi.org/10.1177/0022427810365906.

Caplan, J. M., L. W. Kennedy, J. D. Barnum, and E. L. Piza. 2015. "Risk Terrain Modeling for Spatial Risk Assessment." *Cityscape* 17 (1): 7–15.

Caplan, J. M., L. W. Kennedy, E. L. Piza, and J. D. Barnum. 2020. "Using Vulnerability and Exposure to Improve Robbery Prediction and Target Area Selection." *Applied Spatial Analysis and Policy* 13 (1): 113–136. https://doi.org/10.1007/s12061-019-09294-7.

Chen, T., and C. Guestrin. "XGBoost: A Scalable Tree Boosting System." Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. San Francisco, CA, USA. 13–17 August 2016.

Corcoran, V. 2019. Predicting Violent Crime Reports from Geospatial and Temporal Attributes of US 911 Emergency Call Data.

Cowen, C., E. R. Loiderback, and S. S. Roy. 2019. "The Role of Land Use and Walkability in Predicting Crime Patterns: A Spatiotemporal Analysis of Miami-Dade County Neighborhoods, 2007-2015." *Security Journal* 32 (3): 264–286. https://doi.org/10.1057/s41284-018-00161-7.

Cullen, J. B., and S. D. Levitt. 1999. "Crime, Urban Flight, and the Consequences for Cities." *The Review of Economics and Statistics* 81 (2): 159–169. https://www.jstor.org/stable/2646853.

Detotto, C., and M. Vannini. 2010. "Counting the Cost of Crime in Italy." *Global Crime* 11 (4): 421–435. https://doi.org/10.1080/17440572.2010.519523.

Farrar, D. E., and R. R. Glauber. 1967. "Multicollinearity in Regression Analysis: The Problem Revisited." *The Review of Economic and Statistics* 49 (1): 92–107. https://doi.org/10.2307/1937887.

Garnier, S., J. M. Caplan, and L. W. Kennedy. 2018. "Predicting Dynamical Crime Distribution from Environmental and Social Influences." *Frontiers in Applied Mathematics and Statistics* 4 (13). https://doi.org/10.3389/fams.2018.00013.

Gilstad-Hayden, K., L. R. Wallace, A. Carroll-Scott, S. R. Meyer, S. Barbo, C. Murphy-Dunning, and J. R. Ickovics. 2015. "Research Note: Greater Tree Canopy Cover is Associated with Lower Rates of Both Violent and Property Crime in New Haven, CT." *Landscape and Urban Planning* 143:248–253. https://doi.org/10.1016/j.landurbplan.2015.08.005.

Giménez-Santana, A., J. E. Medina-Sarmiento, and F. Miró-Llinares. 2018. "Risk Terrain Modeling for Road Safety: Identifying Crash-Related Environmental Factors in the Province of Cádiz, Spain." *European Journal on Criminal Policy and Research* 24 (4): 451–467. https://doi.org/10.1007/s10610-018-9398-x.

He, Q., and J. Li. 2021. "The Roles of Built Environment and Social Disadvantage on the Geography of Property Crime." *Cities* 121:1–14. https://doi.org/10.1016/j.cities.2021.103471.

Hipp, J. R., S. Lee, D. Ki, and J. H. Kim. 2022. "Measuring the Built Environment with Google Street View and Machine Learning: Consequences for Crime on Street Segments." *Journal of Quantitative Criminology* 38 (3): 537–565. https://doi.org/10.1007/s10940-021-09506-9.

Jang, H. Y., K. Kim, and J. Y. Lee. 2014. "A Study on the Improvement of CCTV Location for Crime Prevention by citizens' Daily Activity Pattern." *Journal of the Korean Urban Geographical Society* 17 (1): 101–112 (In Korean.

Jang, Y. J., and S. Lee. 2022. "Analysis of the Nonlinear Relationship Between Urban Environment and Fear of Crime: Focusing on CPTED Elements and Interpretable Machine Learning Model." *Journal of the Urban Design Institute of Korea* 23 (3): 142–162. https://doi.org/10.1016/j.scs.2023.104419.

Johnson, S. D., and K. J. Bowers. 2010. "Permeability and Burglary Risk: Are Cul-de-Sacs Safer?" *Journal of Quantitative Criminology* 26 (1): 89–111. https://doi.org/10.1007/s10940-009-9084-8.

Kang, H. W., H. B. Kang, and K.-K. R. Choo. 2017. "Prediction of Crime Occurrence from Multi-Modal Data Using Deep Learning." *PLoS One* 12 (4): e0176244. https://doi.org/10.1371/journal.pone.0176244.

Kim, H. W., D. McCarty, and M. Jeong. 2023. "Examining Commercial Crime Call Determinants in Alley Commercial Districts Before and After COVID-19: A Machine Learning-Based SHAP Approach." *Applied Sciences* 13 (21): 11714. https://doi.org/10.3390/app132111714.

Kim, S., and S. Lee. 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. https://doi.org/10.1016/j.scs.2023.104419.

Kim, Y. A., and J. R. Hipp. 2018. "Physical Boundaries and City Boundaries: Consequences for Crime Patterns on Street Segments?" *Crime & Delinquency* 64 (2): 227–254. https://doi.org/10.1177/0011128716687756.

Lan, M., L. Liu, and J. E. Eck. 2021. "A Spatial Analytical Approach to Assess the Impact of a Casino on Crime: An Example of JACK Casino in Downtown Cincinnati." *Cities* 111:Article 103003. https://doi.org/10.1016/j.cities.2020.103003.

Lee, S. J., and S. J. Kang. 2012. "A Study on the Methodology of Positioning Security CCTV Cameras in Urban Residential District Through Using Space Syntax." *Journal of the Architectural Institute of Korea Planning & Design* 28 (9): 55–62. (In Korean). https://doi.org/10.5659/JAIK_PD.2012.28.9.55.

Lin, J., Q. Wang, and B. Huang. 2021. "Street Trees and Crime: What Characteristics of Trees and Streetscapes Matter." *Urban Forestry & Urban Greening* 65:1–11. https://doi.org/10.1016/j.ufug.2021.127366.

Lipton, R., X. A. Yang, A. Braga, J. Goldstick, M. Newton, and M. Rura. 2013. "The Geography of Violence, Alcohol Outlets, and Drug Arrests in Boston." *American Journal of Public Health* 103 (4): 657–664. https://doi.org/10.2105/AJPH.2012.300927.

Marchment, Z., and P. Gill. 2021. "Systematic Review and Meta-Analysis of Risk Terrain Modelling (RTM) as a Spatial Forecasting Method." *Crime Science* 10 (12): 1–11. https://doi.org/10.1186/s40163-021-00149-6.

Maruthaveeran, S., and C. K. van den Bosch. 2015. "Fear of Crime in Urban Parks–What the Residents of Kuala Lumpur Have to Say?" *Urban Forestry & Urban Greening* 14 (3): 702–713. https://doi.org/10.1016/j.ufug.2015.05.012.

Meng, Y., N. Yang, Z. Qian, and G. Zhang. 2021. "What Makes an Online Review More Helpful: An Interpretation Framework Using XGBoost and SHAP Values." *Journal of Theoretical & Applied Electronic Commerce Research* 16 (3): 466–490. https://doi.org/10.3390/jtaer16030029.

Newman, O. 1972. *Defensible Space: Crime Prevention Through Urban Design*. New York, NY, USA: Macmillan.

Park, C. H., and S. H. Choi. 2009. "Crime Prevention Effects of Publicity of CCTV Installation at Kang-Nam Gu, Seoul: The Effects of First News." *Korean Criminological Review, Korean Institute of Criminology* 79:213–238. (In Korean). https://doi.org/10.5659/JAIK_PD.2012.28.9.55.

Rummens, A., W. Hardyns, and L. Pauwels. 2017. "The Use of Predictive Analysis in Spatiotemporal Crime Forecasting: Building and Testing a Model in an Urban Context." *Applied Geography* 86:255–261. https://doi.org/10.1016/j.apgeog.2017.06.011.

Sampson, R. J. 1983. "Structural Density and Criminal Victimization." *Criminology* 21 (2): 276–293. https://doi.org/10.1111/j.1745-9125.1983.tb00262.x.

Schmid, C. F. 1960a. "Urban Crime Areas: Part I." *American Sociological Review* 25 (4): 527–542. https://doi.org/10.2307/2092937.

Schmid, C. F. 1960b. "Urban Crime Areas: Part II." *American Sociological Review* 25 (5): 655–678. https://doi.org/10.2307/2090139.

Schwartz, A. E., S. Susin, and I. Voicu. 2003. "Has Falling Crime Driven New York City's Real Estate Boom?" *Journal of Housing Research* 14 (1): 101–136.

Schwarz, K., and K. Seidensticker. 2023. "Using Risk Terrain Modeling for the Risk Assessment of Explosive ATM attacks." *Proceedings of The 9th International Conference on Time Series and Forecasting. Engineering Proceedings* 39:24. Gran Canaria, Spain. https://doi.org/10.3390/engproc2023039024.

Sherman, L. W., P. R. Gartin, and M. E. Buerger. 1989. "Hot Spots of Predatory Crime: Routine Activities and the Criminology of Place." *Criminology* 27 (1): 27–56. https://doi.org/10.1111/j.1745-9125.1989.tb00862.x.

Skogan, W. G. 2004. "Community Policing: Common Impediments to Success: The Past, Present and Future." In *Community Policing: The Past, Present and Future 159-167*, Washington, DC: The Annie E. Casey Foundation.

Slack, D., S. Hilgard, E. Jia, S. Singh, and H. Lakkaraju. 2020. "Fooling Lime and Shap: Adversarial Attacks on Post Hoc Explanation Methods." *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 180–186. New York, NY, USA. Association for Computing Machinery.

Sypion-Dutkowska, N., and M. Leitner. 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. https://doi.org/10.3390/ijgi6030074.

Tabangin, D. R., J. C. Flores, and N. F. Emperador. 2008. "Investigating Crime Hotspot Places and Their Implication to Urban Environmental Design: A Geographic Visualization and Data Mining Approach." *International Journal of Civil and Architectural Engineering* 2 (12): 4004–4012.

Tengbeh, S. 2006. *Crime Analysis and Police Station Location in Swaziland: A Case Study in Manzini* (Doctoral dissertation, Stellenbosch: University of Stellenbosch).

Theobald, D. M., D. L. Stevens, D. White, N. S. Urquhart, A. R. Olsen, and J. B. Norman. 2007. "Using GIS to Generate Spatially Balanced Random Survey Designs for Natural Resource Applications." *Environmental Management* 40 (1): 134–146. https://doi.org/10.1007/s00267-005-0199-x.

Thomas, S. A., C. T. Harris, and G. Drawve. 2022. "Exploring the Influence of Elements of the Social and Physical Environment on Neighborhood Gun Crime." *American*

Journal of Criminal Justice 47 (3): 370–398. https://doi.org/10.1007/s12103-020-09599-1.

Tong, X., P. Ni, Q. Li, Q. Yuan, J. Liu, H. Lu, and G. Li. 2021. "Urban Crime Trends Analysis and Occurrence Possibility Prediction Based on Light Gradient Boosting Machine." *Proceedings of the 2021 IEEE 4th International Conference on Big Data and Artificial Intelligence*, 98–103. Qingdao, China.

Troy, A., J. M. Grove, and J. O'Neil-Dunne. 2012. "The Relationship Between Tree Canopy and Crime Rates Across an Urban–Rural Gradient in the Greater Baltimore Region." *Landscape and Urban Planning* 106 (3): 262–270. https://doi.org/10.1016/j.landurbplan.2012.03.010.

Twinam, T. 2017. "Danger Zone: Land Use and the Geography of Neighborhood Crime." *Journal of Urban Economics* 100:104–119. https://doi.org/10.1016/j.jue.2017.05.006.

Vildosola, D., J. Carter, E. R. Louderback, and S. Sen Roy. 2019. "Crime in an Affluent City: Applications of Risk Terrain Modeling for Residential and Vehicle Burglary in Coral Gables." In *Applied Spatial Analysis and Policy*, 441–459. Vol. 13. Florida, 2004–2016. https://doi.org/10.1007/s12061-019-09311-9.

Wang, M., K. Zheng, Y. Yang, and X. Wang. 2020. "An Explainable Machine Learning Framework for Intrusion Detection Systems." *Institute of Electrical and Electronics Engineers Access* 8:73127–73141. https://doi.org/10.1109/ACCESS.2020.2988359.

Weisburd, D., C. White, S. Wire, and D. B. Wilson. 2020. "Enhancing Informal Social Controls to Reduce Crime: Evidence from a Study of Crime Hot Spots." *Prevention Science* 22 (4): 509–522. https://doi.org/10.1007/s11121-020-01194-4.

Wheeler, A. P., and W. Steenbeek. 2020. "Mapping the Risk Terrain for Crime Using Machine Learning." *Journal of Quantitative Criminology* 37:1–36. https://doi.org/10.1007/s10940-020-09457-7.

Witt, R., A. Clarke, and N. Fielding. 1999. "Crime and Economic Activity. A Panel Data Approach." *The British Journal of Criminology* 39 (3): 391–400. https://doi.org/10.1093/bjc/39.3.391.

Wolfe, M. K., and J. Mennis. 2012. "Does Vegetation Encourage or Suppress Urban Crime? Evidence from Philadelphia, PA." *Landscape and Urban Planning* 108 (2–4): 112–122. https://doi.org/10.1016/j.landurbplan.2012.08.006.

Yirmibesoglu, F., and N. Ergun. 2007. "Property and Personal Crime in Istanbul." *European Planning Studies* 15 (3): 37–41. https://doi.org/10.1080/09654310601017067.

Zhang, X., L. Liu, M. Lan, G. Song, L. Xiao, and J. Chen. 2022. "Interpretable Machine Learning Models for Crime Prediction." *Computers, Environment and Urban Systems* 94, Article 101789. 101789. https://doi.org/10.1016/j.compenvurbsys.2022.101789.

Zhang, X., L. Liu, L. Xiao, and J. Ji. 2020. "Comparison of Machine Learning Algorithms for Predicting Crime Hotspots." *Institute of Electrical and Electronics Engineers Access* 8:181302–181310. https://doi.org/10.1109/ACCESS.2020.3028420.