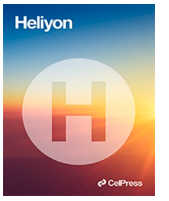
Heliyon 10 (2024) e38405

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Street-level built environment on SARS-CoV-2 transmission: A study of Hong Kong

Chongyang Ren [a,b,1], Xiaoran Huang [a,c,*,1], Qingyao Qiao [b], Marcus White [c]

[a] School of Architecture and Art, North China University of Technology, Beijing, 100144, China
[b] Faculty of Architecture, the University of Hong Kong, Hong Kong
[c] Centre for Design Innovation, Swinburne University of Technology, Hawthorn, Victoria, 3122, Australia

ARTICLE INFO

Keywords:
Street-level built environment
Spatial transmission risk
COVID-19 pandemic
Computer vision
Interpretable machine learning
Hong Kong

ABSTRACT

Understanding the association between SARS-CoV-2 Spatial Transmission Risk (SSTR) and Built Environments (BE) is crucial for implementing effective pandemic prevention measures. Massive efforts have been made to examine the macro-built environment at the regional level, which has neglected the living service areas at the residential scale. Therefore, this study aims to explore the association between Street-level Built Environments (SLBE) and SSTR in Hong Kong from the 1st to the early 5th waves of the pandemic to address this gap. A total of 3693 visited/resided buildings were collected and clustered by spatial autocorrelation, and then Google Street View (GSV) was employed to obtain SLBE features around the buildings. Eventually, the interpretable machine learning framework based on the random forest algorithm (RFA)-based SHapley Additive exPlanations (SHAP) model was proposed to reveal the hidden non-linear association between SSTR and SLBE.

The results indicated that in the high-risk cluster area, street sidewalks, street sanitation facilities, and artificial structures were the primary risk factors positively associated with SSTR, in low-risk cluster areas with a significant positive association with traffic control facilities. Our study elucidates the role of SLBE in COVID-19 transmission, facilitates strategic resource allocation, and guides the optimization of outdoor behavior during pandemics for urban policymakers.

1. Introduction

The ongoing Coronavirus Disease 2019 (COVID-19) pandemic has had profound implications for public health and has significantly impacted various sectors and economies globally since its emergence in December 2019 [1]. As of April 30, 2023, WHO's 114th edition of Weekly Epidemiological Update, over 765 million confirmed cases and more than six million deaths have been recorded worldwide [2]. Previous research indicates that transmission risk within high population density metropolitan areas was heterogeneous and related to various built environment factors [3]. The built environment (BE) was a crucial factor for non-pharmacological measures, and it was feasible to make specific environmental modifications to prevent the spread of viruses, such as improving transportation

* Corresponding author. School of Architecture and Art, North China University of Technology, Beijing, 100144, China.
E-mail addresses: cyren@hku.hk (C. Ren), xiaoran.huang@ncut.edu.cn (X. Huang), qingyaoq@hku.hk (Q. Qiao), marcuswhite@swin.edu.au (M. White).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.heliyon.2024.e38405
Received 20 May 2024; Received in revised form 21 September 2024; Accepted 24 September 2024
Available online 25 September 2024

accessibility and increasing green spaces in densely populated areas; these modifications can effectively influence human behaviors and interactions, ultimately reducing the probability of SARS-CoV-2 transmission risk [4,5].

This study explored the association of overlooked Street-level Built Environments (SLBE) with SARS-CoV-2 Spatial Transmission Risk (SSTR). 3693 buildings involved in 28,003 visited/reside COVID-19 cases records. SLBE features were generated using 84,045 Hong Kong Google Street View (GSV) images, eventually establishing the spatial connection between buildings and outdoor walkable service areas. This spatial relationship evaluation surpassed the constraints of solely relying on SLBE as a covariate in regional-level research [6,7]. To aim to investigate the following three sub-questions.

1) What is the better approach to revealing the association between SBLE and SSTR?
2) What SLBE features are more influential, and are their impacts different across the various waves of the pandemic?
3) From a policy and management perspective, what specific SLBE should be prioritized for spatial management to reduce street-level transmission effectively?

2. Literature reviews

The concept of the BE varies across different scientific fields. Public health refers to external physical conditions, including structures, areas, and items designed or altered by individuals' health [8,9]. These factors significantly impact the movement of people in specific spaces, which in turn affects disease transmission [10]. The revealed risk factors in the BE contributing to SARS-CoV-2 transmission include crowding, poverty, racism reflected in housing and neighborhood characteristics, inadequate indoor air circulation, and ambient air pollution [11]. However, past studies in spatial epidemiology have typically utilized a regional-level scale to investigate the relationship between SSTR and the BE to balance the safeguard of privacy [12,13].

2.1. Necessary to assess the impact of SLBE

Some recent studies suggested that there may be an unrevealed association between the SLBE and SSTR. Regarding spatial factors, Hamidi et al.'s research indicated that traffic connectivity played a more significant role in the spread of the pandemic, primarily in large metropolitan areas with numerous economic, social, and commuting linkages [14]. In terms of transmission probabilities, although indoor transmission is more likely, outdoor transmission cannot be ignored, especially with higher air pollutant levels, leading to increased morbidity and mortality [15]. Another important finding showed that the virulence of SARS-CoV-2 can survive for 3 h in aerosols and up to 72 h on surfaces like plastic and stainless steel in outdoor environments [16]. Furthermore, Because nations had initially implemented a range of restricted policies to address the transmission, a multitude of mental illnesses, including anxiety, depression, post-traumatic stress disorder, and psychological distress, had increased incidence rates varying from 6.33 % to 81.9 % [17]. Thus, leading prevention policies turn into encouraging outdoor exercise for residents to remain mentally active and access essential outdoor space usage and living supplies, indirectly increasing outdoor infectious risk [18]. These recent studies highlight the importance of exploring the association between SLBE and SSTR, which may have been overlooked in assessing SARS-CoV-2 transmission risk.

2.2. Computer vision's potential in SLBE analysis

The historical method for obtaining recognition of SLBE was often through on-site interviews and telephone surveys [19]. On the one hand, these low-efficiency approaches pose challenges in quantification and utilization in studies conducted at the desired spatial and geographic scale [19].In contrast, measurements are time-consuming and expensive but provide limited heuristic insights for policymakers [20]. The use of abstract metrics to describe SLBE is another common approach. A study investigated the spatial arrangement of the street network with the spread of SARS-CoV-2 by analyzing the locations of 3815 confirmed cases in Hong Kong before the 2nd wave, indicated that areas with higher levels of integration (a measure of the cognitive complexity of pedestrians face when navigating a street) and betweenness centrality values (a measure of spatial network accessibility) tend to have more COVID-19 confirmed cases [21]. However, the space syntax questioned does not fully represent SLBE. It is more of an auxiliary method based on the topological method of limits as a representation of simple streets and urban zoning boundaries. Also, it only had the function of describing the evaluation of abstract spatial concept indexes [22,23].

Recent GSV images have proven to be a more efficient method for large-scale objective data collection than costly and time-consuming on-site auditing [24]. It is publicly accessible, capturing street-level images of urban streets throughout metropolitan areas [25]. Meanwhile, GSV image audits produce a high agreement (75–96.7 %) with in-person auditing, and the intra-rater agreements are also high, indicating that this approach is reliable for evaluating SLBE features [26]. The GSV segments at the beginning were used in non-communicable diseases (NCD) within public health research to analyze the inherent street physical features and their association with chronic diseases such as diabetes and obesity in communities [27]. GSV have demonstrated 78 % and 80 % confidence in the validity and reliability of investigating chronic NCD, respectively [28]. Research has shown that certain street physical features, such as crosswalks, are associated with a lower ratio of obesity and premature mortality, and the presence of visible wires and single-lane highways has been linked to higher obesity rates and reduced physical activity [29,30]. After the COVID-19 pandemic, an analysis of 164 million GSV images across the United States on a zip code scale found that mixed land use, sidewalks, dilapidated buildings and visible wires were associated with higher COVID-19 cases, green streets are the opposite [6].

2.3. Existing evaluating implements

The optimal assessment method for assessing the association between BE and SSTR remains controversial. Geographically Weighted Regression (GWR) model was widely adopted for BE investigation in several countries and states, including Brazil, Indonesia, and Africa [31–34], and it was more accurate than other models, including the Long Short-Term Memory (LSTM) and Susceptible-Exposed-Infected-Recovered-Dead (SEIRD) model [35]. But GWR accounting for the spatial location effect in its results, it remained fundamentally a linear model, which was less effective in revealing the nonlinear association between BE and SSTR. As researchers gradually began exploring the nonlinear effects of BE, they increasingly employed machine learning (ML) models [36,37]. Among these, methods that combined spatial autocorrelation with ML were particularly valuable, as they considered both the spatial location effect and nonlinear associations in their evaluations [36].

3. Materials and method

The comprehensive process of this study is presented in Fig. 1 and consists of three main phases. (1) Data collection: Download the 1st to the early 5th visit and residence buildings of COVID-19-confirmed cases from open data sources. Also, the GSV image was downloaded and segmented into different SLBE variables. (2) Data processing: The spatial autocorrelation analysis on the SSTR was conducted, considering both Tertiary Planning Units (TPU) and building-weighted risk. Subsequently, SLBE variables of the 800-m walking service range were matched to a COVID-19 related building. (3) Interpretable machine learning: The random forest algorithm (RFA)-based SHapley Additive exPlanations (SHAP) model was used to interpret the overall ranking of values, interactions between street physical features, and variations among successive outbreak waves.

3.1. Study area

Hong Kong is a special administrative region in China's eastern Pearl River Delta. It is comprised of 18 administrative districts and 291 TPUs. The city features a well-developed BE with compactness and connectivity, with public transportation accounting for 88.8 % of the overall road network [38]. Since the first SARS-CoV-2 confirmed imported case from Wuhan on January 23, 2022, Hong Kong has experienced five waves of COVID-19 outbreak involving Alpha, Delta, and Omicron variants. As of April 10, 2023, 9148 deaths and 748,050 confirmed cases were recorded (https://www.coronavirus.gov.hk/sim/index.html).

Despite a busy transportation hub and high population density, Hong Kong did not impose a city lockdown. Instead, it implemented a series of dynamic strict measures before the 1st wave till the early part of the 5th wave, depending on the severity of the pandemic, including social quarantine, telecommuting, traffic control measures, school closures, bar and restaurant closures, and official telecommuting, which limited the scope of residents' lives and reduced the number of transmissions [39,40]. This resulted in a decrease in public transport and local travel by 36.1 % and 52.3 %, and an increase of 32.2 % in indoor time compared with 2019 [41,42].

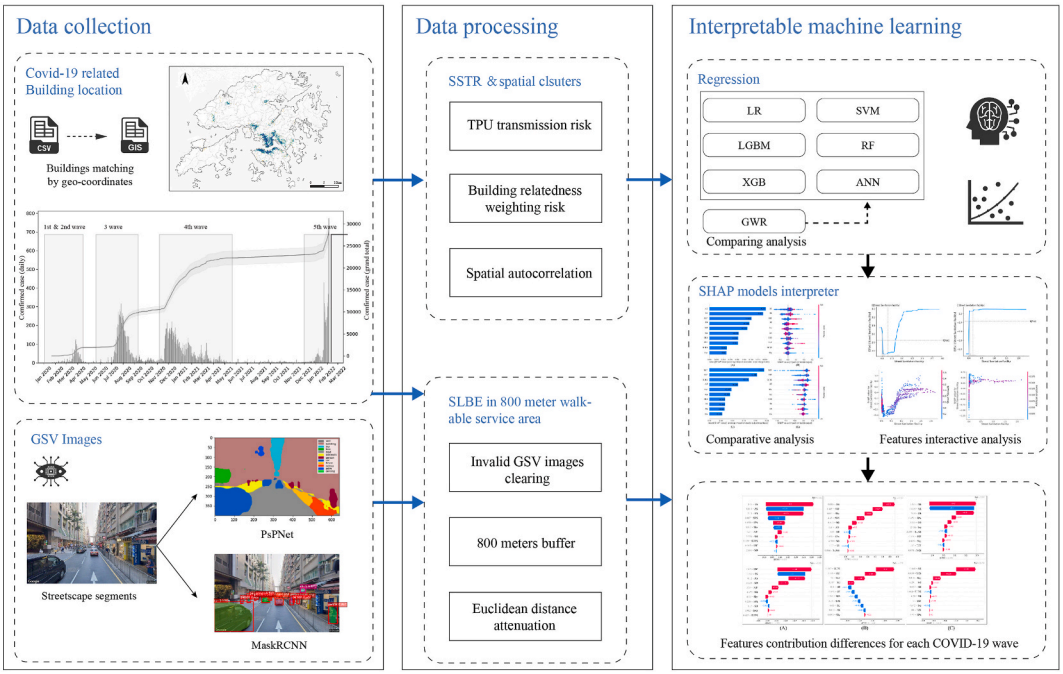


Fig. 1. The schematic framework for assessing the association between SSTR and SLBE.

Residents were more likely to relocate within proximity of their homes. The travel-restricted regulations in Hong Kong provided an ideal setting for studying the association between SLBE around COVID-19 related buildings and SSTR during different pandemic waves.

3.2. COVID-19 cases location

Hong Kong Common Spatial Data Infrastructure Portal offered data access to the cumulative COVID-19 confirmed cases (https://portal.csdi.gov.hk/geoportal/#metadataInfoPanel), which includes confirmed dates and building geo-coordinates where confirmed cases resided and visited. Fig. 2 shows the spatial distribution of buildings (n = 3693) related to COVID-19-confirmed cases during the five waves. Table 1 summarizes the number of confirmed cases and involved buildings during each pandemic wave.

3.3. Independent variables

Due to the limited space of the main body and for the sake of concision, the statistical description of independent variables was

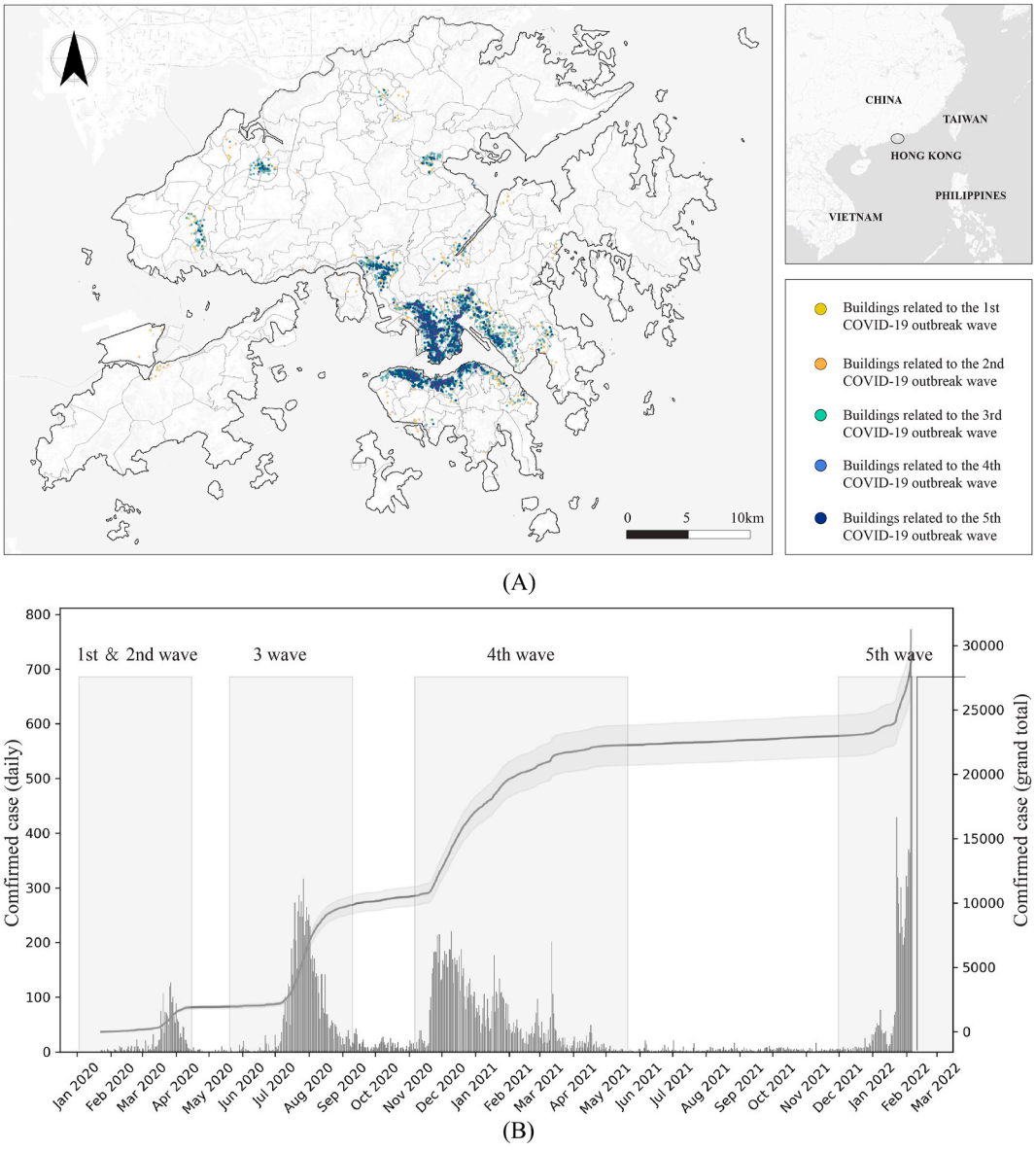


Fig. 2. Spatial-temporal distribution of Hong Kong COVID-19 outbreak: (A) The spatial distribution of COVID-19-related buildings, (B) The temporal accumulation of COVID-19 confirmed cases.

Table 1
The number of COVID-19-related cases and related buildings.

COVID-19 outbreak waves	Number of visited/reside records	Number of related buildings
1st wave (Jan–Mar 2020)	1435	607
2nd wave (Apr–Jun 2020)	694	123
3rd wave (Jul–Sept 2020)	7996	1724
4th wave (Nov 2020–May 2021)	11802	931
5th wave (Jan–Mar 2022)	6076	308

summarized in Appendix 1.

3.3.1. SSTR weighted risk

The regional-level SSTR was defined as the proportion of cases to population and spatial risk density based on TPU's administrative areas [43,44]. The Hong Kong TPUs, consisting of 291 units, were obtained TPU boundaries from the Hong Kong Common Spatial Data Infrastructure Portal as well (https://portal.csdi.gov.hk/geoportal/#metadataInfoPanel), and then we integrate the building-level transmission risk by weighting the number of confirmed cases of residents and visits at the corresponding building. Subsequent references to SSTR were all at the building level, and its mathematical expression was described in Equation (1)

$$SSTR = \frac{\omega^b \bullet N_c}{S \bullet N_p} \tag{1}$$

Where.

N_c = Number of COVID-19 confirmed resident and visit cases within a single TPU
N_p = Total population within the TPU
S = Area of the TPU in square kilometres
ω^b = Normalized building weight risk coefficient for the number of COVID-19 confirmed resident and visit cases, ranging from 0 to 1

3.3.2. SLBE features

3.3.2.1. GSV images. The road network data was downloaded from the Data.Gov.HK(https://data.gov.hk/en-data/dataset/hk-landsd-openmap-3d-pedestrian-network), which covers the Hong Kong primary and secondary roads, significant footways inside villages, major footpaths inside country parks and the footways inside estates. As illustrated in Fig. 3(A), the Sampling frequency along the pedestrian network was set at 50-m intervals, representing a balance between the minimum image collection required and the maximum coverage of the entire pedestrian network by referring to the past method [45–47]. The GSV images (n = 84,045) were acquired by Google API and named by Place IDs (PID). The highest image resolution available was 800 × 400 pixels, with a camera pitch of heading angle forward (0°) and a horizontal field of view of 120° to simulate a pedestrian's visual field. 84,045 PIDs were downloaded geographically across Hong Kong in Fig. 3(B).

3.3.2.2. GSV images classification and physical feature segmentation. We employed two pixel-level recognition algorithms to segment SLBE features for GSV images. The first was the pyramid scene parsing network (PSPNet), known for its accurate semantic segmentation performance in street visual elements proportion [48]. Another segmentation quantity method, the Mask R-CNN, was calculated for traffic feature quantities [49,50]. Prior research had used Street Greenness (SG), Street Building (SB), street sidewalks (SS), Street Road Surface (SRU), Street Unpaved Road Surface (SURS), Traffic Control facilities (TCF), and sky as classification categories for SLBE

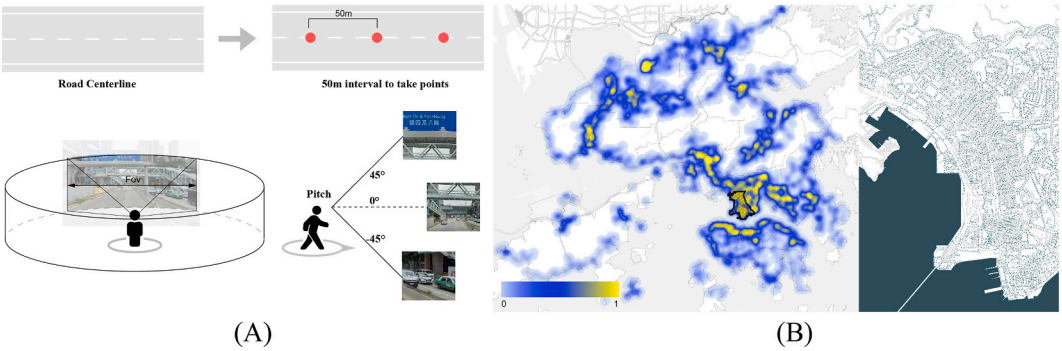


Fig. 3. GSV image capture approach: (A) Simulation of GSV image perspective, (B) Spatial density distribution of FIDs.

[6,7]. We consulted earlier research on the streetscape integration approach, which had the potential to lead to more widely applicable policy recommendations [51], and using building density replaces building type [43]. We categorized objects into Street Sanitation Facilities (SSF), Artificial Structures (AS), Street Obstacles (SO), and Street Public Amenities (SPA). Fig. 4 is an example of streetscape segmentation, including prominent elements such as walls, trees, sidewalks, buildings, roads, earth, sky, trucks, buses, cars, and persons. Eventually, the street's physical features were classified into 10 main classes with 26 specific elements, as outlined in Table 2.

3.4. Method

3.4.1. 800-Meter walkable service area

We correlated the COVID-19-related buildings with segmentation results of the GSV image of 800-m walkable service areas. The 800-m walkable service area was frequently employed in urban travel behavior studies to examine residents' walking range [52,53]. To better simulate the residents' exposed frequency on the street and to account for Hong Kong's challenging topography and its implications for road planning, we employed Euclidean distances to assign weights to the GSV image segmentation results. This approach was commonly used in studies related to the transmission of infectious diseases and geographical accessibility [54].

Fig. 5 demonstrated the integration of GSV images with COVID-19-related buildings to determine the value of each GSV image in calculating SLBE. This calculation was based on the Euclidean distance exponentially decaying locations within an 800-m walkable service. The calculation formula was described in Equation (2):

$$GSV_i = \exp(-\alpha * D(B_i, G_i)) \tag{2}$$

Where.

B_i = COVID-19 confirmed case related building location
G_i = GSV image i
$D(B_i, G_i)$ = Euclidean distance between B_i and G_i.

3.4.2. Multi-collinearity estimation

The variance inflation factor (VIF) and tolerance (also known as 1/VIF) are closely related statistics for diagnosing multi-

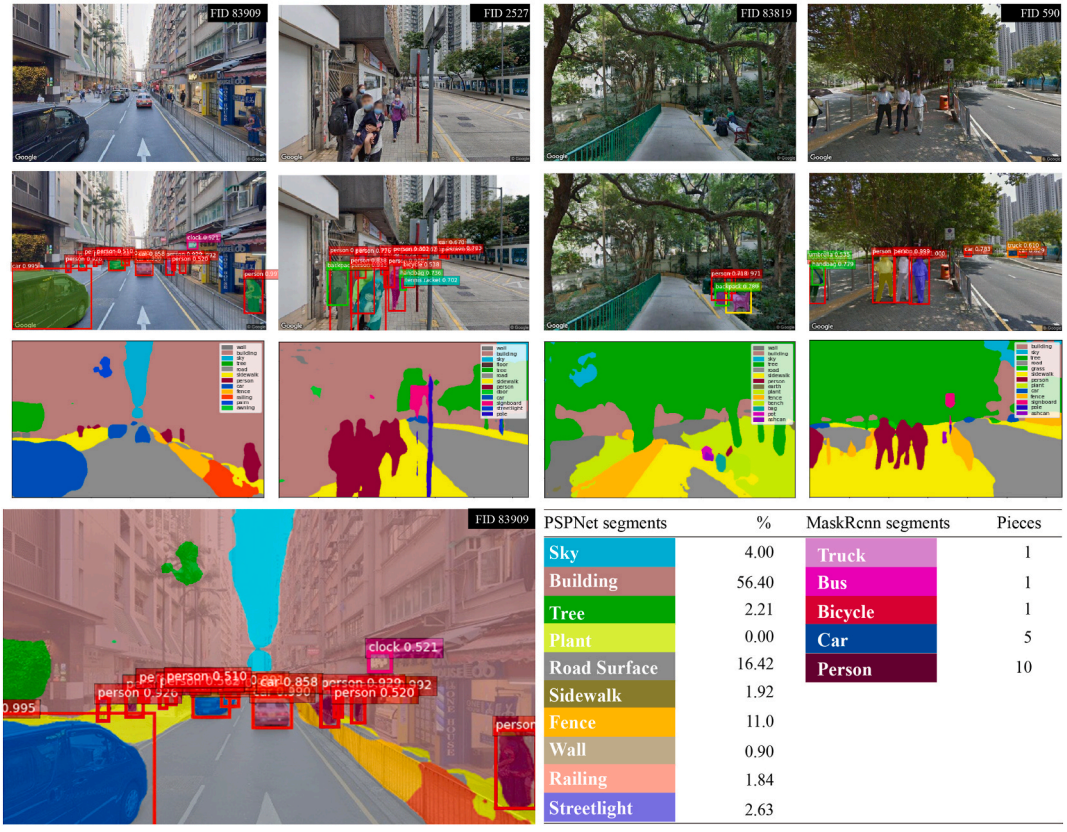


Fig. 4. Streetscape segmentation by PSPNet and Mask R-CNN.

Table 2
Street physical features classes and sub-street attributes.

Street physical feature classes	Sub-street elements	Description
Street greenness	P_tree	The percentage of tree cover (%)
	P_grass	The percentage of grass areas (%)
	P_plant	The percentage of plant areas (%)
	P_green_hill	The percentage of green hill areas (%)
Street building	P_skyscraper	The percentage of skyscrapers (%)
	P_windowpane	The percentage of window glass (%)
	P_normal_building	The percentage of standard buildings (excluding skyscrapers) (%)
	P_glass	The percentage of glass curtain walls (%)
Street sidewalk	P_sidewalk	The percentage of sidewalks (pedestrian walkways) (%)
Street road surface	P_road_surface	The percentage of paved road surfaces (e.g., asphalt, concrete) for motorized and non-motorized vehicles (%)
Street unpaved road surface	P_vacant_land	The percentage of bare road surfaces (e.g., wasteland, exposed ground)
Street sanitation facilities	P_trash_can	The percentage of waste disposal facilities available (%)
Artificial facilities	P_stall	The percentage of street vendor stalls or kiosks (%)
	P_awning	The percentage of awnings or overhead coverings (%)
	P_billboard	The percentage of advertising billboards or large promotional displays (%)
	P_signboard	The percentage of informational signboards or directional signs (%)
Street obstacles	P_low_wall	The percentage of low walls hindering movement (%)
	P_fence	The percentage of fences restricting access (%)
	P_railing	The percentage of railings providing safety or boundaries (%)
Street public amenities	P_sculpture	The percentage of sculptures or statues (%)
	P_fountain	The percentage of fountains features (%)
	P_bench	The percentage of bench seating (%)
	P_streetlight	The percentage of outdoor streetlights (%)
Sky	P_sky	The percentage of the visible sky (%)
Traffic control facilities	N_stop_sign	The number of stop signs
	N_traffic_light	The number of traffic lights

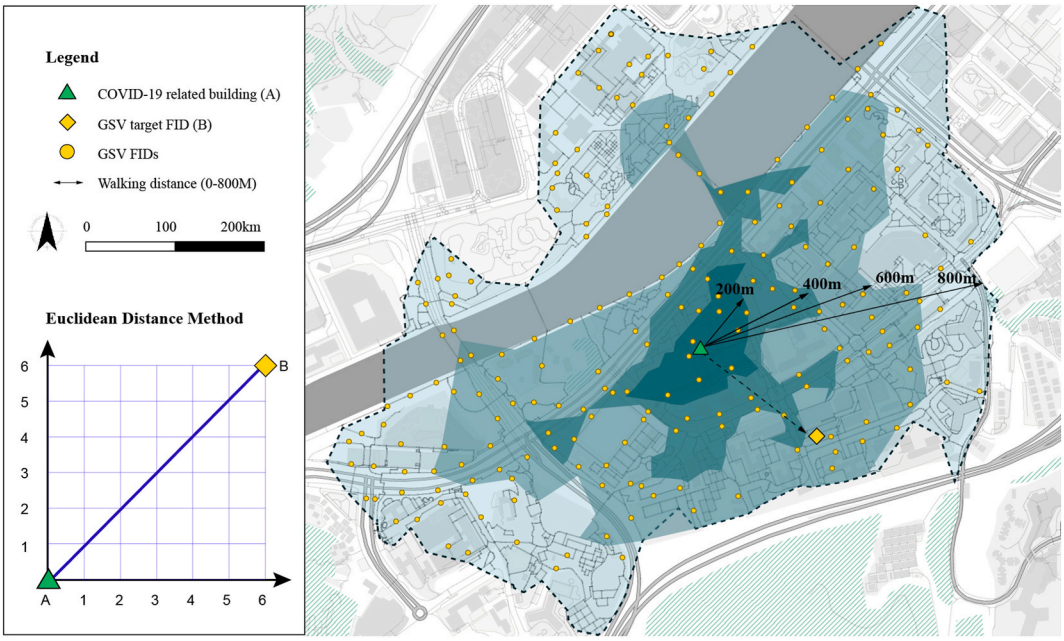


Fig. 5. GSV images weight calculation by 800-m walkable service areas.

collinearity in multiple regression [55]. The former measures the degree of multi-collinearity between an independent variable and all other independent variables, while the latter, being the reciprocal of VIF, indicates the independence level of a variable from the other independent variables [56]. They are based on the R-squared (R^2) score obtained by regressing a predictor on all the other predictors in the analysis. A rule of thumb frequently given is that the 1/VIF should not be less than 0.1, and consequently, VIF should not be greater than 10 [57].

As illustrated in Table 3, multi-collinearity was assessed before further analysis by estimating VIF and 1/VIF for the HH and LL

clusters across the 1st to 5th waves. Overall, the LL cluster parameters exhibited better control than the HH cluster, with all VIF values less than 5 and all 1/VIF values greater than 0.2. However, in the HH cluster, SPA and SB had VIF values greater than 10 and 1/VIF values less than 0.1, indicating potential significant multi-collinearity issues between these variables. Our strategy was to remove SPA for two reasons: first, this adjustment reduced all VIF values to below 9 and increased all 1/VIF values to above 0.1, providing better control than removing SB. SPA also contributed larger multi-collinearity values in the LL cluster. Its removal would also help improve the reliability of both clusters.

3.4.3. Spatial clustering

Spatial autocorrelation is a method that examines the similarity or correlation between neighbouring units in a geographic area. It was commonly employed to investigate the spatial clustering distribution of epidemic situations [58,59]. This paper utilized global spatial correlation, specifically the global Moran's I score, to assess whether spatial autocorrelation effects existed in COVID-19-related buildings. Moran's I score near 1 indicates positive spatial autocorrelation, values near 0 indicate no significant spatial autocorrelation, and values near −1 indicate negative spatial autocorrelation. Also, the P-value measures significance, and the Z-score shows the significance and direction of clustering trends; typically, a P-value below 0.05 is considered statistically significant, and a Z-score above 1.96 or below −1.96 is considered significant. As the results are indicated in Table 4, the spatial autocorrelation was evident across all outbreak waves and most pronounced when measuring the completed 1st-5th waves.

The Global Moran's I index unable to adjust for the stability of local spatial processes. Therefore, it is necessary to examining the local spatial correlation characteristics of the COVID-19 related buildings. Local Indicators of Spatial Association (LISA) cluster map is a common approach to assistant visualization [60,61]. As demonstrated in Fig. 6 (A), the buildings were classified into High-High (HH) clusters, which indicate regions with aggregation of associated high-case-related buildings, and Low-Low (LL) clusters, which indicate regions with aggregation of low-case-related buildings. High-Low (HL) outliers also represented regions with high-case-related buildings associated with low-case-related buildings. Similarly, Low-High (LH) outliers represented regions with low-case-related buildings associated with high-case-related buildings, and Not-Significant categories meant a lack of significant aggregation effect.

Fig. 6 (B) illustrates each category's number and spatial distribution. The outlier and non-significant categories were excluded from the following analysis due to their non-uniform distribution and the lack of potential for discussing generalized results. The result was consistent with the finding that the local spatial cluster categories in the Chinese mainland primarily comprised the HH and LL clusters during the initial transmission of SARS-CoV-2 [60].

3.4.4. Interpretable machine learning

We compared six typical supervised ML regression models to reveal the association between STTR and SLBE [37,62], including.

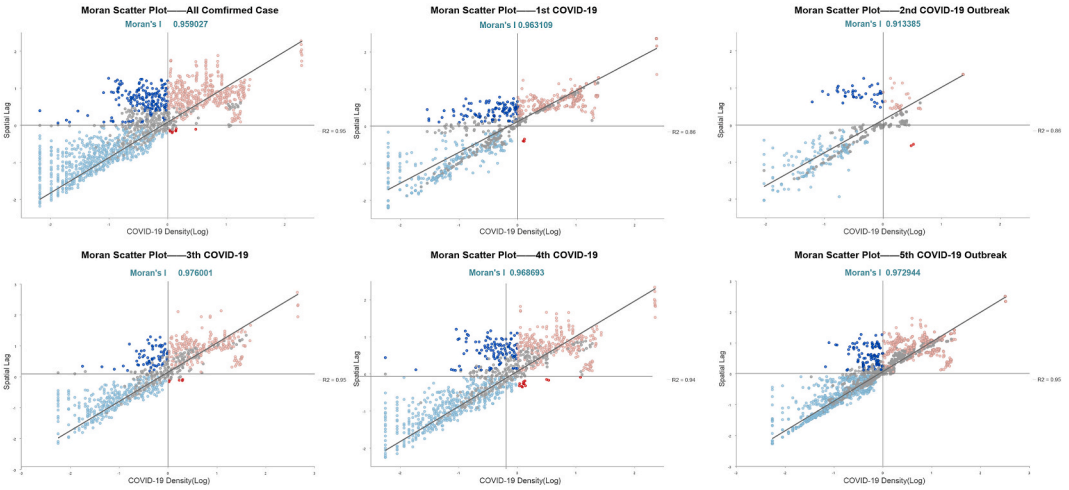
- Linear regression (LR) is a statistical technique that establishes a mathematical model that describes the connection between input features and the target. The objective is to minimize the linear approximation by minimizing the sum of squared discrepancies between the actual values and the predicted values.
- Support Vector Machine (SVM) is a binary classification/regression model that operates based on the principle of hyperplane separation. This approach guarantees the identification of the hyperplane that may efficiently divide the training datasets within the most significant geometric interval by utilizing a Kernel function to facilitate mapping input spaces onto a feature space with many dimensions.
- Light Gradient Boosting Machine (LGBM) is a boosting-based methodology that demonstrates superior prediction speed and accuracy performance compared to alternative boosting and bagging algorithms. The method under consideration employs a gradient-boosting decision tree algorithm, which integrates gradient-based one-sided sampling with exclusive feature-bundling techniques. LGBM differs from the traditional way of dividing trees in gradient boosting machines by utilizing a leafwise technique to improve accuracy by creating more complex models.
- XGBoost (XGBT) is based on decision trees and utilizes a gradient-boosting framework. This framework enables parallel tree boosting and combines the outcomes of many Classification and Regression Trees into an ensemble.

Table 3
Multi-collinearity estimation.

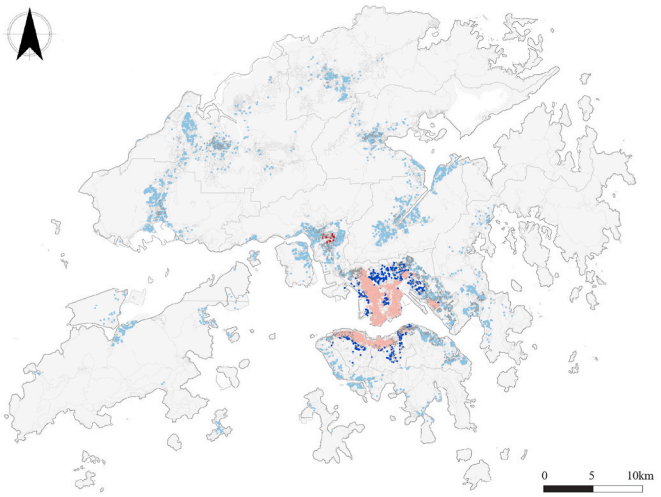
HH cluster					LL cluster	
Variables	VIF	1/VIF	VIF (without SPA)	1/VIF (without SPA)	VIF	1/VIF
SG	5.79	0.17	5.79	0.17	1.75	0.57
SSF	3.08	0.32	2.72	0.36	1.73	0.57
SPA	12.45	0.08	/	/	4.27	0.23
SB	16.42	0.06	8.92	0.11	2.38	0.41
SO	3.30	0.30	3.28	0.30	2.26	0.44
SS	8.26	0.12	8.23	0.12	2.29	0.43
SRS	9.032	0.111	8.528	0.117	4.282	0.234
SURS	5.434	0.184	5.368	0.186	3.863	0.259
TCF	2.306	0.434	2.296	0.435	1.701	0.588
AS	3.367	0.297	3.334	0.300	1.233	0.811
Sky	1.925	0.519	1.745	0.573	2.385	0.419

Table 4
Global Moran's I index of the COVID-19 related buildings from 1st to 5th waves.

Metrics	1st wave	2nd wave	3rd wave	4th wave	5th wave	1st-5th wave
Moran's I	0.70	0.62	0.42	0.32	0.41	0.74
P-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Z-Score	49.89	12.18	124.90	60.78	20.52	439.34



(A)



High-High Cluster
High-Low Outiler
Low-High Outiler
Low-Low Cluster
Not Significant

	HH	HL	LH	LL	NS
Buildings related to the 1st wave	228	0	177	31	171
Buildings related to the 2nd wave	64	0	4	19	36
Buildings related to the 3rd wave	965	4	120	197	438
Buildings related to the 4th wave	637	11	73	61	149
Buildings related to the 5th wave	240	3	24	14	27

(B)

Fig. 6. A) Global Moran scatter plot integrated with LISA results, (B) LISA spatial distribution and clustering quantities across various waves.

- Artificial neural networks (ANN) replicate the structural arrangement of human neurons, comprising three interconnected layers of neurons: the input layer, the hidden layer, and the output layer. In ANN, the neurons inside each layer connect with neurons in the next and previous layers using a weighted summation function. The input data is transmitted from the input layer to the output layer to calculate the prediction error. The error is subsequently backpropagated to adjust the weights of the neurons.
- Random forest (RF) is a widely recognized ensemble technique that combines a predetermined number of decision trees. The criterion for distributing each node among all trees is either information gain or Gini impurity. In general, it is observed that the initial nodes in decision trees tend to display the highest degree of impurity reduction. In contrast, the last nodes exhibit the lowest degree of impurity reduction. Therefore, creating a subset that includes the most essential features is feasible by intentionally eliminating branches at a specific node.

We utilized grid search for model hyperparameter selection with datasets division of training (75 %) and testing (25 %). Grid search

is a systematic method that optimizes model performance by exhaustively exploring predefined hyperparameter combinations, ensuring finding the optimal solution within a specific parameter space and enhancing model accuracy and robustness [63]. Since we used the grid search with the combined results from 1st to 5th, it was able to reflect the overall characteristics of the epidemic fluctuations and obtain a more representative and stable choice of parameters.

The outcome is displayed in Table 5. For the LR model, default settings were adopted due to the relatively limited range of parameter tuning, which is due to default settings already providing sufficient benchmark performance. Extensive hyperparameter tuning was performed for tree-based models, including RF, LGBM, and XGB. Key parameters tuned included the number of estimators (n_estimators), which reflects the number of base learners in the ensemble, and the minimum samples split (min_samples_split), which specifies the minimum number of samples required to split a node, helping to prevent overfitting. Additionally, the maximum depth (max_depth) controls the maximum depth of the trees, and the minimum samples per leaf (min_samples_leaf) indicates the minimum number of samples required to be at a leaf node, both contributing to model regularization and minimizing overfitting. For the boosting tree models LGBM and XGB, the learning rate (learning_rate) was tuned, determining the step size at each iteration while moving toward a minimum loss function. Although a lower learning rate requires more training time, it often results in better convergence to the optimal solution. The subsample and colsample_bytree parameters, which controlled the proportion of data and features used per tree, respectively, were critical for preventing overfitting. The C and gamma parameters were adjusted for the SVM model. The C parameter controls the trade-off between achieving a low error on the training data and minimizing the norm of the weights, while the gamma parameter defines the influence of a single training example. Lastly, the Adam optimizer was employed for the ANN, and the number of epochs was set for training. The network architecture included two hidden layers, each with 32 neurons and ReLU activation functions, and a single neuron output layer for regression tasks.

As ML results evaluation, we used the R^2 score as a statistical measure to evaluate regression models' fitness performance and Root Mean Square Error (RMSE), defined as the standard deviation of prediction errors [64,65].

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - p_i)^2} \tag{3}$$

$$R^2 = \frac{n\left(\sum_{i=1}^{n} y_i p_i\right) - \left(\sum_{i=1}^{n} y_i\right)\left(\sum_{i=1}^{n} p_i\right)}{\sqrt{\left[n\left(\sum_{i=1}^{n} {y_i}^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right)\right]\left[n\left(\sum_{i=1}^{n} {p_i}^2 - \left(\sum_{i=1}^{n} p_i\right)^2\right)\right]}} \tag{4}$$

where y_i is the ground truth and p_i is the predicted value.

The SHAP model is an interpretable ML Interpretive method based on game theory, which provides a different explanation pathway base on feature contribution compared to traditional statistical methods [66]. It calculates the Shapley value corresponding to each variable using the Monte Carlo sampling method [67]. SHAP was embedded with ML and had been widely employed to improve medical decision-making accuracy [68] and detect health risk factors [69].Ths calculation is defined in Equation (5) (6):

$$g(z') = \phi_0 + \sum_{i=1}^{M} \phi_i {z_i}' \tag{5}$$

Where $z' \in \{0,1\}^M$ represents whether a feature is used to estimate the output variable, M is the number of input features, ϕ_i is the SHAP value of the ith feature, and ϕ_0 is the mean value of the output variable. The SHAP value assesses feature importance by comparing model prediction performance with and without each SLBE feature in feature combinations:

$$\phi_i = \sum_{S \subseteq z' \{i\}} \frac{|S|!(M - |S| - 1)}{M!}[f_x(S \cup \{i\}) - f_x(S)] \tag{6}$$

Where S is the set of non-zero z', and $f_x(S) = E[|f(x)x_S|]$ is the expected outcome of the model $f(x)$ subjected to S.

3.4.5. GWR comparison

The study also compared the common GWR method in past studies [31–34]. Specifically, the purpose was to evaluate the differences in model performance, prediction accuracy, and the ability to capture spatial heterogeneity when revealing the association

Table 5
Parameter settings of various models in HH and LL clusters.

	HH cluster	LL cluster
LR	default	default
SVM	{'C': 9.796, 'gamma': 'auto', 'kernel': 'rbf'}	'C': 9.494, 'gamma': 'auto', 'kernel': 'rbf'
LGBM	{'learning_rate': 0.087, 'max_depth': 7, 'n_estimators': 300}	learning_rate': 0.070, 'max_depth': 7, 'n_estimators': 300
RF	{'n_estimators': 300, 'min_samples_split': 2, 'min_samples_leaf': 1, 'max_features': 'sqrt', 'max_depth': None}	n_estimators': 300, 'min_samples_split': 2, 'min_samples_leaf': 1, 'max_features': 'sqrt', 'max_depth': None
XGB	'colsample_bytree': 0.865, 'learning_rate': 0.099, 'max_depth': 7, 'n_estimators': 100, 'subsample': 0.810	colsample_bytree': 0.743, 'learning_rate': 0.073, 'max_depth': 7, 'n_estimators': 150, 'subsample': 0.790
ANN	optimizer = 'adam', epochs = 150, Dense (32, activation = 'relu'), Dense (32, activation = 'relu'), Dense (1)	optimizer = 'adam', epochs = 100, Dense (32, activation = 'relu'), Dense (32, activation = 'relu'), Dense (1)

between SLBE and SSTR. This comparison can highlight which method suits specific datasets and research objectives. The R^2 metric is commonly used to evaluate the fitness performance of GWR models, like ML models [70].

However, R^2 is not an adequate evaluation criterion for GWR fitness, AIC(Akaike information criterion), BIS(Bayesian information criterion), RSS(Residual sum of squares) as a supplementary evaluation method, a smaller AIC value represents a better GWR model and the degree of fit [71]. The AIC evaluates model quality by considering the model's fit and complexity. Lower absolute values indicate a model that better balances fit and simplicity. The BIC is a model selection criterion like the AIC, while the BIC imposes a more stringent penalty for model complexity. The RSS metric measures the difference between predicted and observed values. Smaller RSS values indicate better model fit due to reduced prediction error.

4. Results and findings

4.1. Performance of ML

The fitting performance and prediction errors of various ML models on the training and testing datasets are evaluated in Tables 6 and 7, respectively. The former reflects the result of the HH cluster, while the latter was about the LL cluster. The SLBE exhibits an obvious non-linear association with STTR, as evidenced by the significant differences in R^2 between the LR model and other non-linear models. The fitting for the HH cluster surpasses that of the LL cluster across all models, likely due to sample size issues, which led to unsuccessful fitting in the 2nd and 5th waves of the LL cluster.

The LGBM and RF models achieved consistently high R^2 and RMSE values across the various waves on both clusters' training and test sets. However, RF is slightly better. Specifically, RF has shown slightly better fitting performance for each wave in the HH cluster while exhibiting greater stability in the LL cluster. In contrast, the R^2 for LGBM significantly decreased to a mere 0.07 in the 4th wave in the LL cluster. We further visualized the RF model's training and testing consistency performance in Fig. 7 to assess its fit across different data points. The results indicate that the model performs well overall in fitting the HH and LL cluster trends despite the deficiencies in capturing the extremes.

4.2. Performance of GWR

In addition, the performance results of the GWR were listed in Table 8, and the R^2 for each wave ranged from 0.75 to 0.94, which indicates that the model effectively captures the primary patterns of the data for each pandemic wave. However, the overall R^2 value in the 1st-5th wave was only 0.74, contrary to the expectation that a larger sample size would result in better model fit. Thus, observing the AIC and BIC values, which dramatically increased for all waves except the 1st and 2nd waves, suggests that the model may be overfitting and experiencing larger errors in local spatial regions.

Thus, we spatially visualized RSS and R^2 values over the 291 TPUs for the 1st-5th wave, as shown in Fig. 8 (A&B). It was noticed that the GWR performed better in predicting SSTR suburban and peri-urban areas and performed poorly in urban areas. The inability of the GWR to predict the SSTR in high-density urban areas, where many buildings associated with COVID-19 confirmed cases are located, made RF the most suitable method for SHAP analysis.

Table 6
ML performance estimation in the HH cluster.

HH cluster-training												
	1st wave		2nd wave		3rd wave		4th wave		5th wave		1st-5th wave	
Method	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
LR	0.55	0.75	0.24	1.24	0.32	1.02	0.40	0.91	0.37	0.98	0.32	1.03
SVM	0.83	0.45	0.68	0.80	0.73	0.64	0.74	0.59	0.83	0.50	0.74	0.63
LGBM	0.98	0.12	0.42	1.08	0.98	0.14	0.97	0.16	0.98	0.16	0.98	0.16
RF	0.96	0.22	0.85	0.54	0.97	0.23	0.96	0.22	0.95	0.26	0.96	0.17
XGB	0.99	0.02	0.99	0.02	0.99	0.07	0.99	0.04	0.99	0.02	0.99	0.10
ANN	0.75	0.57	0.52	0.99	0.78	0.54	0.74	0.53	0.77	0.58	0.86	0.46
HH cluster-testing												
	1st wave		2nd wave		3rd wave		4th wave		5th wave		1st-5th wave	
Method	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
LR	0.24	1.17	0.31	0.87	0.17	1.19	0.25	1.03	0.06	1.37	0.28	1.04
SVM	0.25	1.10	0.48	0.75	0.60	0.82	0.59	0.77	0.44	1.06	0.70	0.67
LGBM	0.27	1.10	0.36	0.84	0.80	0.58	0.59	0.76	0.34	1.15	0.87	0.42
RF	0.28	1.10	0.50	0.74	0.82	0.54	0.68	0.67	0.37	1.12	0.86	0.45
XGB	0.23	1.12	0.29	0.88	0.82	0.55	0.70	0.64	0.39	1.10	0.85	0.47
ANN	0.16	1.17	0.46	0.76	0.69	0.73	0.59	0.72	0.42	1.07	0.80	0.54

Table 7
ML performance estimation in LL cluster.

LL cluster-training												
	1st wave		2nd wave		3rd wave		4th wave		5th wave		1st-5th wave	
Method	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
LR	0.49	0.91	/	/	0.35	0.75	0.20	0.66	/	/	0.40	0.93
SVM	0.88	0.40	/	/	0.82	0.39	0.94	0.17	/	/	0.89	0.39
LGBM	0.96	0.25	/	/	0.94	0.21	0.43	0.55	/	/	0.98	0.16
RF	0.95	0.27	/	/	0.93	0.23	0.88	0.24	/	/	0.96	0.23
XGB	0.99	0.01	/	/	0.99	0.01	0.99	0.01	/	/	0.99	0.04
ANN	0.70	0.70	/	/	0.77	0.74	0.80	0.32	/	/	0.86	0.46
LL cluster-testing												
	1st wave		2nd wave		3rd wave		4th wave		5th wave		1st-5th wave	
Method	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
LR	0.28	1.31	/	/	0.12	0.79	/	/	/	/	0.37	1.00
SVM	0.38	1.12	/	/	0.46	0.62	0.14	0.69	/	/	0.62	0.79
LGBM	0.54	1.04	/	/	0.54	0.57	0.07	0.72	/	/	0.67	0.75
RF	0.49	1.11	/	/	0.50	0.60	0.48	0.53	/	/	0.66	0.74
XGB	0.45	1.14	/	/	0.54	0.57	0.51	0.52	/	/	0.68	0.71
ANN	0.50	1.08	/	/	0.30	0.71	/	/	/	/	0.63	0.77

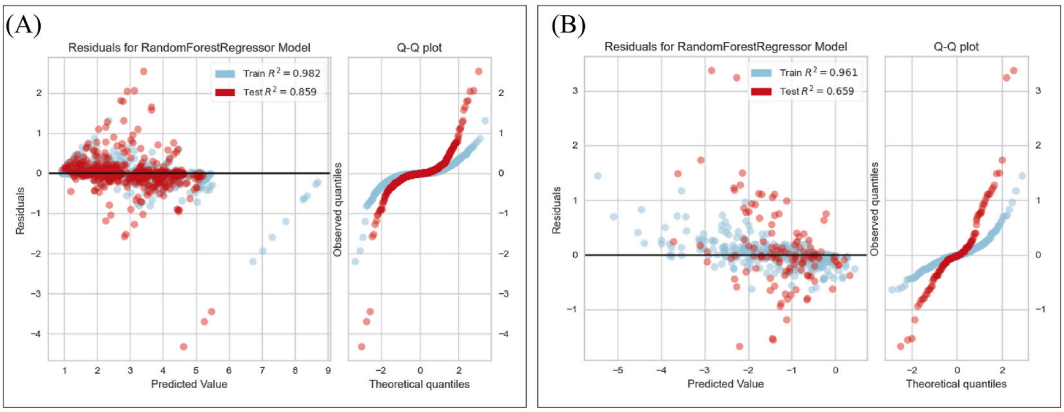


Fig. 7. RF fitness consistency analysis:(A) HH cluster residuals and R^2 in 1st-5th wave, (B) LL cluster residuals, and R^2 in 1st-5th wave.

Table 8
GWR model fitting effect comparison between five-wave COVID-19 outbreaks.

Metrics	1st wave	2nd wave	3rd wave	4th wave	5th wave	1st-5th wave
R^2	0.93	0.75	0.94	0.87	0.76	0.74
AIC	249.66	205.07	277.57	874.86	558.60	−1313.94
BIC	989.29	293.51	2888.63	2003.39	802.45	5761.22
RSS	30.85	22.08	71.12	84.51	53.84	86.94

4.3. RFA-SHAP model analysis

The SHAP approach was utilized to enhance the interpretability of machine learning models and examine the impact of input features on the model's output. Fig. 9 presents an overview of the SHAP summary plot to analyze the association and importance of street physical features and SSTR for the 1st-5th wave. The x-axis units represent the RF model's change in average SHAP value impact on model output magnitude units of SSTR in Fig. 9(A & C). The distribution of SHAP values for each feature is provided by overlapping points scattered vertically, and the SHAP x-value represents the effect of a feature on the SSTR prediction model for that feature. Each colored dot in the plot corresponds to the magnitude of the variable's impact on the SSTR results, as indicated by the gradient scale on the x-axis in Fig. 9 (B&D).

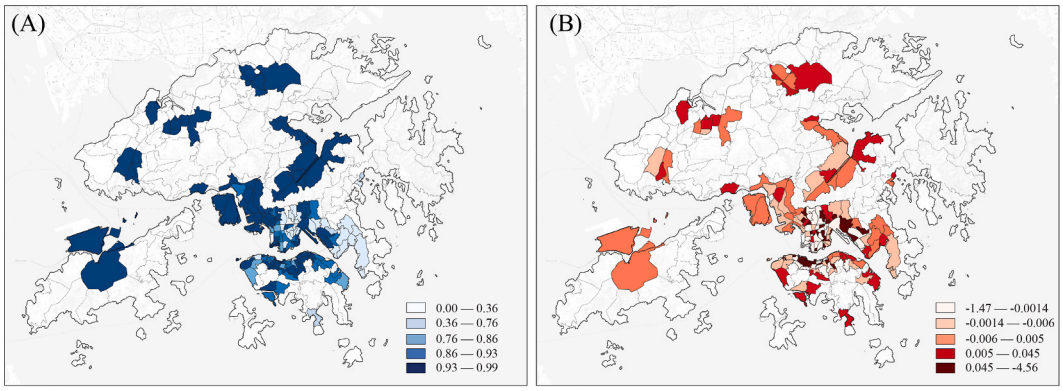


Fig. 8. (A) GWR R^2 spatial distribution from 1st to 5th wave on TPUs, (B) GWR RSS spatial distribution from 1st to 5th wave on TPUs.

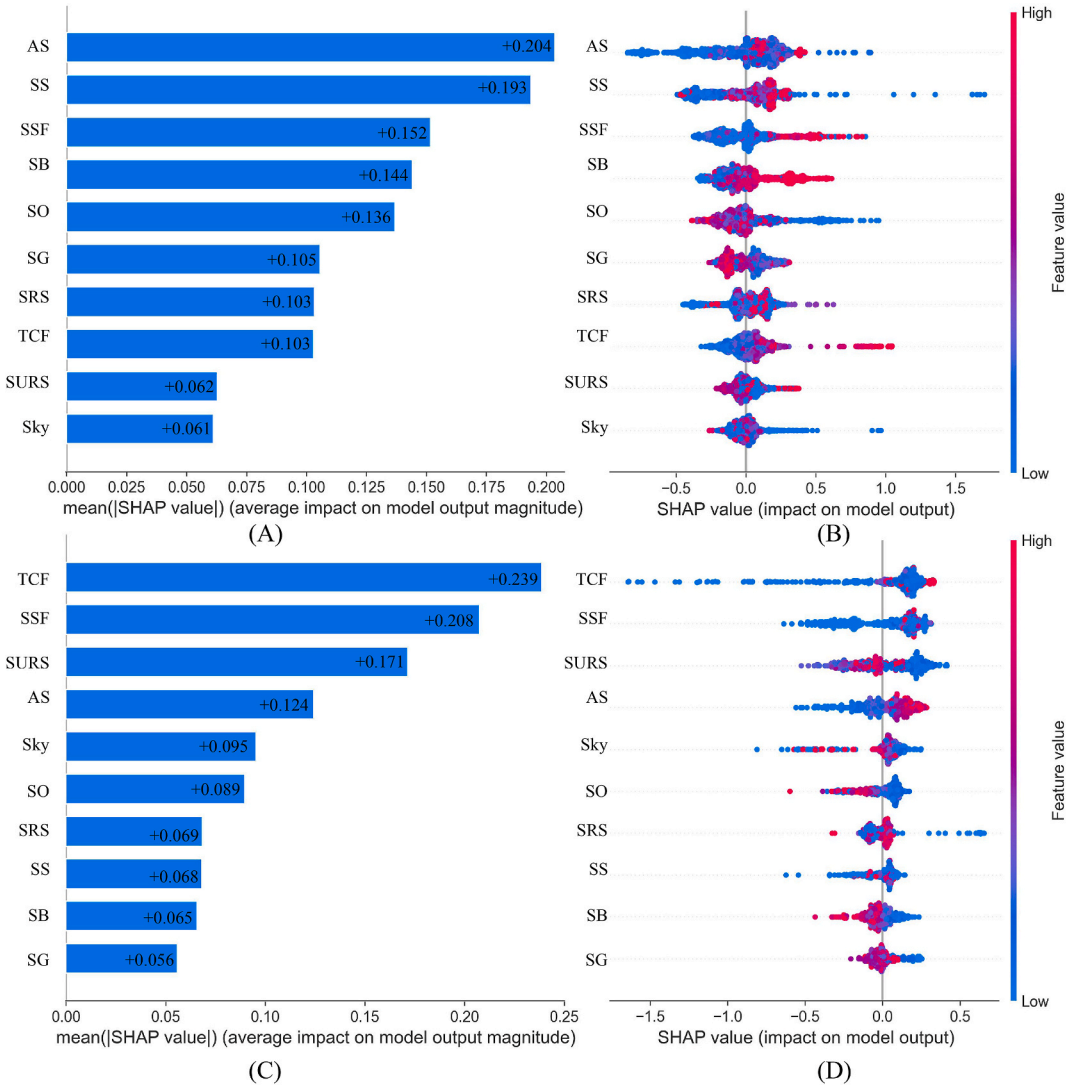


Fig. 9. SHAP importance and correlation value in 1st to 5th waves; (A) Mean SHAP value importance in HH cluster; (B) SHAP value correlation in HH cluster; (C) Mean SHAP value importance in LL cluster; (D) SHAP value correlation in LL cluster.

1) Regarding street physical features in the HH cluster, the SHAP value of AS, SS, SSF, SB, and SO exceeds 0.13, as shown in Fig. 9(A), indicating their greater significance compared to other variables. Aside from SO, these factors positively correlated with the model output in Fig. 9(B). However, it is worth noting that the association of those top influent variables tends to be a nonlinear pattern, except SB has an obvious positive association.
2) In Fig. 9(C), a SHAP value of 0.10 differentiates the high SHAP value contributor of the LL cluster. Besides SURS having a negative association with SHAP values, the TCF, SSF, and AS demonstrated non-linearity pattern in Fig. 9(D).

Thus, we generated partial correlation plots to analyze the primary contributing features of the two clusters and gain a deeper understanding of the nonlinear relationships. It is evident that these primary contributing characteristics are positively associated with the HH cluster (Fig. 10(A–D))and negatively associated with the LL cluster (Fig. (E to H)). Notably, their feature contribution is not gradual but exhibits a sudden change followed by a stable region.

For features among the HH cluster. AS shown in Fig. 10 (A), the risk suddenly increases when the density exceeds 0.05 %, after which the SHAP values stabilize around 3.2, but with strong downward fluctuations in the syphilis density (0.05 %–0.15 %) range. Fig.10 (B) demonstrates a notable and gradual decline in the positive association between SS and SSTR. The positive SHAP value decreases by nearly half (from 5.0 to around 2.7) at SS density(~0.75 %), indicating a higher likelihood of outdoor transmission in areas with narrower SS density, and the SHAP value exhibited greater instability and susceptibility to multiple factors. As shown in Fig. 10(C), SSF is quite unique. It experiences an initial decrease followed by stabilization, and then the SHAP value increases again from 2.8 to 3.3 at a density greater than approximately 0.5 %. As depicted in Fig. 10(D), a small proportion of SO, less than 4 %, demonstrated an increased risk of transmission, but beyond this range, the SHAP value decreased fluctuatedly.

In the LL cluster, as shown in Fig. 10(E–G), SSF, TCF, and AS all exhibit a diminishing negative correlation as their proportions rise. Among them, the increase in AS is highly volatile, suggesting potential influence from other variables. Conversely, a higher density of SURS could lead to a dramatic reduction in SSTR, with 0.8 % being a distinct threshold, unlike SO, which does not show significant fluctuation in reduction. Additionally, the SSF results in both clusters reveal that SSF demonstrates a robust positive correlation within the HH cluster but an opposite correlation and clear sudden SHAP value change in the LL cluster. The reason for this discrepancy requires further revealing.

We further examine the interactive effects using interactive SHAP maps for those fluctuating primary contributing features shown (Fig. (11)), including AS, SS, and SSF in both clusters. Because it is possible that the genuine impacts were caused in conjunction with other primary contributing features, this also helps us interpret the non-linear fluctuations in their effects.

In the HH cluster, as shown in Fig. 11(A). The higher distribution of AS and SSF at medium SS density (>1.5 %) may counteract each other, particularly at low SS density (~0.5 %), potentially leading to a negative effect on SSTR. Also, a negative correlation is especially apparent in SO. Regarding the interacting effects on SSF, which mainly reveal the decreased slope range (from 0 to 1.0 %), are associated with higher AS and SO values, as depicted in Fig. 11(B). The most crucial influence among them is that of SO. In the LL cluster, Fig. 11(C) demonstrates that SURS consistently exhibits a negative correlation, and the sudden SHAP value increase is associated with AS and TCF. Specifically with extremes of TCF density. In Fig. 11(D), higher SS density appears to interact with the AS rising SHAP process (ranging from 0.05 % to 0.20 %). However, it is unclear how the interactions impact the other two features.

Additionally, we analyze the feature contribution of SLBE variables in the 1st,3rd, and 4th waves, which are effectively fitted by the ML model ($R2 > 0.5$). By using the average SSTR of different waves as the baseline value E[f(x)] and f(x) as the predicted outcome after adding SLBE variables, we can clearly observe the positive or negative contributions of these variables in each wave. As shown in Fig. 12 (A & B). In all clusters across different waves, f(x) is greater than E[f(x)], suggesting that these variables contribute positively to the predictions after accounting for the SLBE variables, leading to an increased model prediction of COVID-19 transmission risk. In the

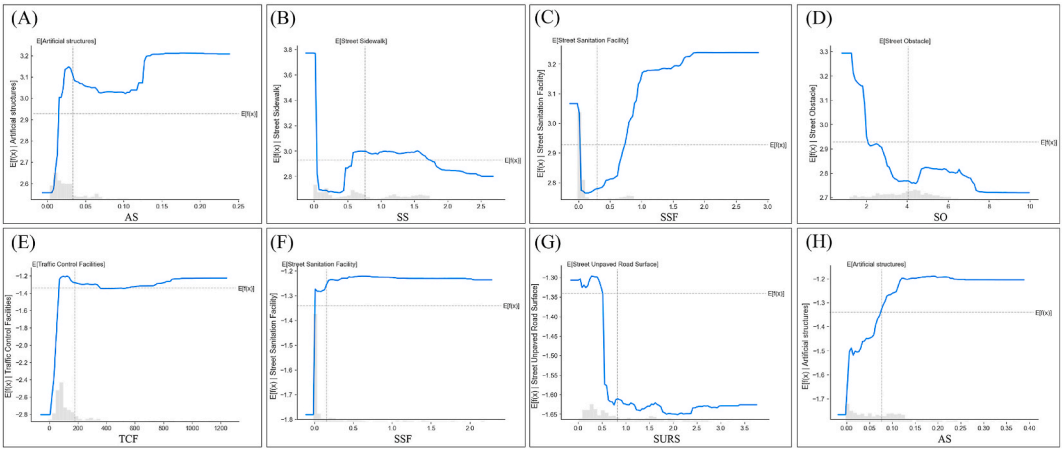


Fig. 10. Main contributed features partial correlation: (A) AS partial correlation in the HH cluster, (B) SS partial correlation in the HH cluster, (C) SSF partial correlation in the HH cluster, (D) SO partial correlation in the HH cluster, (E) TCF partial correlation in the LL cluster, (F) SSF partial correlation in the LL cluster, (G) SURS partial correlation in the LL cluster, (H) AS partial correlation in the LL cluster.

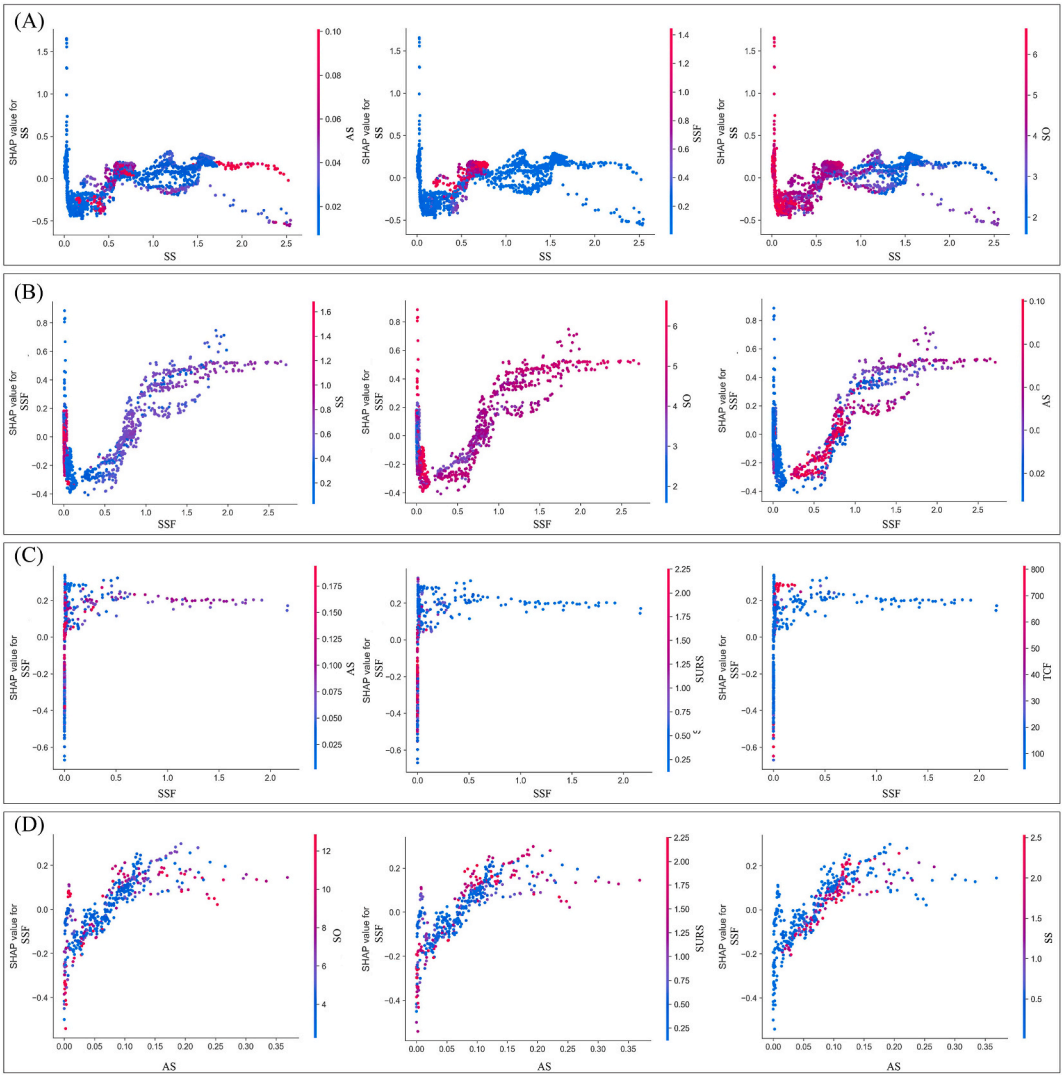


Fig. 11. Features: (A) SS interaction in HH cluster analysis, (B) SSF interaction in HH cluster analysis, (C) SSF interaction in LL cluster analysis, (D) AS interaction in LL cluster analysis.

HH cluster, SB, AS, and SS, are the main contributors in 1st and 4th waves, But SS exhibits inconsistent effects. During the 3rd wave, street infrastructure elements like SSF and AS become more positively pronounced. SS and SB are less significant in the LL clusters, with SSF and AS standing out as the primary positive contributors. Additionally, street natural features (e.g., SURS and SG) gradually influence SSTR.

5. Discussion

5.1. Main findings

This study made the following four research highlights.

1) The results show RF model can perform better in explaining the nonlinear relationship between SLBE and SSTR than other ML and GWR model in building-related 800-Meter walkable service area. Additionally, this study is the first to use the RFA-SHAP interpretable model to analyze associated Hong Kong SLBE features with SSTR from the 1st to early 5th wave of the COVID-19 outbreaks.
2) This study is innovatively based on the buildings where the diagnosed cases visited and resided. It combined the building-related 800-m walkable service with spatial autocorrelation to define SLBE under restricted travel conditions.

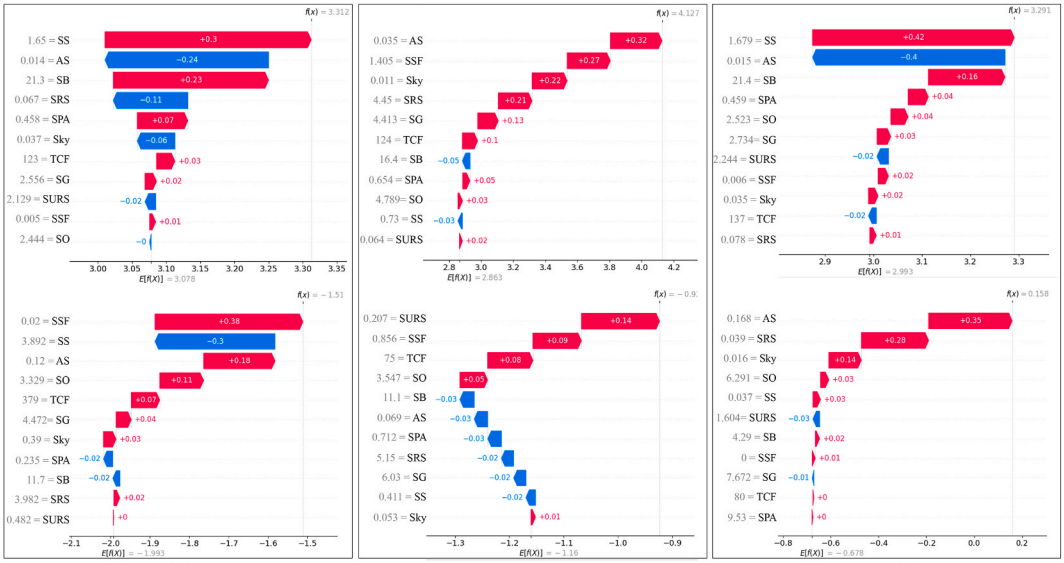


Fig. 12. (A) SLBE-variables value in the 1st wave; (B) SLBE-variables value in the 3rd wave; (C) SLBE-variables value in the 4th wave.

3) This study also highlights the importance of street physical features in influencing SSTR. We identified consistent results with prior studies. For example, residents of areas with higher SB and SS density had higher severity of COVID-19 transmissions. In contrast, higher levels of SG helped reduce the transmission risk, consistent with previous studies [6,12]. The only difference was in the SB within the LL cluster, but since it ranked very low in contribution, we did not focus on it. Instead, several interesting street risk factors and spatial patterns of SSTR are often overlooked but are essential in preventing respiratory transmissions and promoting health equity. In the HH cluster area (such as Sheung Wan and Tsim Sha Tsui) with a high-density area, we found that street infrastructure (e.g., AS, SO, and SSF) correlated highly with transmission risk. A similar distribution was observed in the LL cluster area (such as Tuen Mun and Tsuen Wan) with a low COVID-19 confirmed cases distribution, but the presence of TCF and SSF was highly contributory. Outdoor vacant land (named SURS in the study) was mainly unconstructive land where the reduction in SSTR was greater than SG. Regarding the synergistic effects of SLBE characteristics, we found that (a) in the HH cluster region, a reasonable distribution of SO contributes to the reduction of SSTR with low to medium density of SS or SSF, and (b) in the LL cluster region, the co-presence of TCF and SSF is risker for SSTR, whereas the presence of SO in areas with more AS is also effective in reducing risk as well.
4) In addition, we also analyzed the impact contribution of the three waves of outbreaks in both clusters. We found that street infrastructure and hard space (e.g., SS, SSF, AS) and street natural features (e.g., Sky, SG, SURS) revealed high contribution in the 1st wave of the outbreak with minor population diagnoses situation and in large-scale COVID-19 outbreaks in the 3rd and 4th waves.

5.2. *Mechanisms and policy suggestions*

Based on the above results, the following three recommendations for future infectious disease prevention regarding policy implications.

1) Some causal relationships between solid waste exposure and health outcomes have been demonstrated, but others remain unclear or have not been prioritized as a current public health issue [72]. In particular, people working in waste collection are more likely to be severely affected by injuries, respiratory complications, and transmissions due to their lack of protective equipment [73,74]. Prior studies have shown that indirect contact with surfaces contaminated with infectious droplets can lead to SARS-CoV-2 contamination through waste from COVID-19-affected patients or residents undergoing isolation treatment in areas other than private homes, hospitals, and healthcare centres [75]. Therefore, not only do we need to reduce the spread of SARS-CoV-2 through household waste generated by COVID-19 patients or isolates through proper packaging, incineration, and well-managed landfills. Our study suggests that we also need to clean SSFs that confirmed populations might come into contact with when they are out and about, as pedestrians may come into contact with these solid waste recycling facilities and thus carry infectious toxins on their hands, increasing community morbidity [76]. Meanwhile, prioritizing the management of SSF around TCF may be of more value.

To better clarify policy guiding recommendations, we will further discuss the average SHAP value of the SSF spatial distribution in Hong Kong's 18 districts. Fig. 13 displays the distribution of the mean SHAP values of SSF in the HH and LL clusters. Even in densely

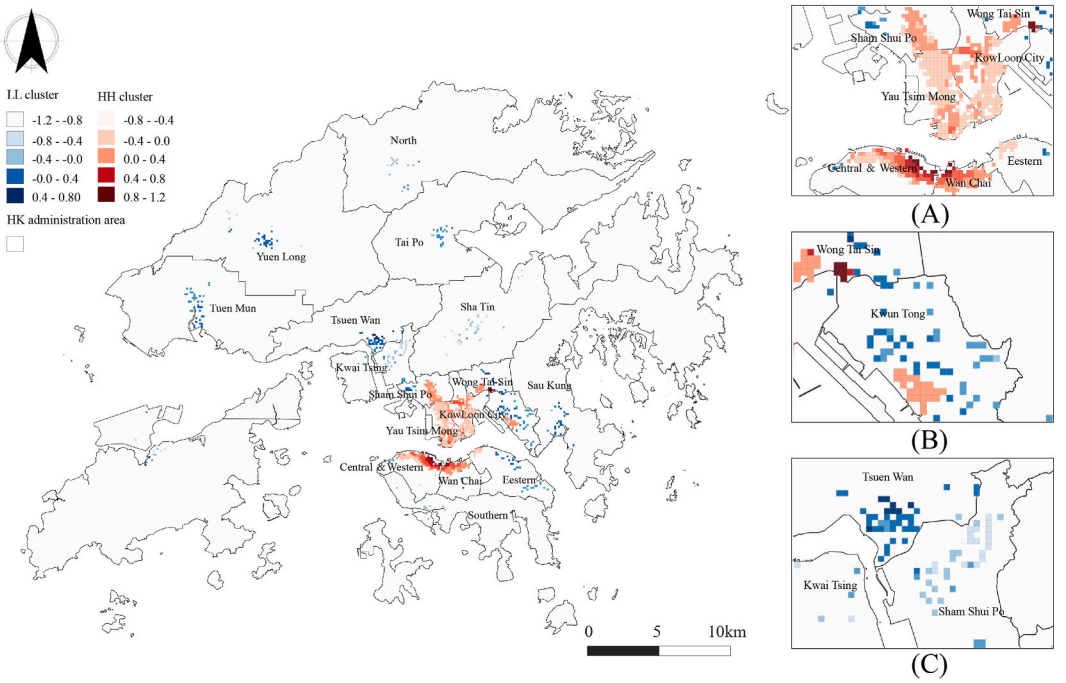


Fig. 13. Spatial distribution of mean SHAP values of the SSF: (A)Mean SHAP value of SSF at HH cluster area, (B) Mean SHAP value of SSF at mixed cluster area, and (C) Mean SHAP value of SSF at LL cluster area.

populated and urbanized parts of Hong Kong, the distribution of the HH cluster does not show a clear positive or negative correlation. The regions with a strong positive correlation (0.0–1.2) are primarily located in the southern part of Yau Tsim Mong and Kowloon City. The areas with a high positive correlation (0.4–1.2) are concentrated in Central & Western and Wan Chai, as seen in Fig. 13(A). Within the mixed cluster area depicted in Fig. 13(B), there is no notable disparity in the positive and negative correlations observed in the average SHAP values. The LL cluster has a concentration of locations with a high positive correlation (0.4–0.8) around the confluence of Tsuen Wan and Sham Shui Po, as depicted in Fig. 13(C).

2) SO is another notable street physical feature that deserves careful consideration, while it did not significantly contribute to the RFA-SHAP model of overall five-wave or individual wave studies. However, its interacting effects on high-risk SS and AS, SSF could effectively reduce SSTR. They are considering that AS.SSF is disseminated through SS; we hypothesize that SO may cause people to maintain a certain physical distance in SS. Prior studies confirmed that maintaining a physical distance of 1 m or more outdoors can significantly reduce person-to-person transmission of SARS-CoV-2 [77].Thus, erecting barriers to restrict people's movement can reduce the rate of SARS-CoV-2 transmission [78]. The results of this article may indicate that arranging and implementing physical barriers to reduce mobility and physical contact, such as physical collisions in some higher-risk areas outdoors, mainly it is vital to consider when the SSF and SS are less than 1.5 % in the HH cluster and when the AS less 1.5 % in the LL cluster.
3) Closed public recreational facilities and a ban on public gatherings can reduce COVID-19 transmission through built environment-mediated pathways [79]. However, these studies focus on closed indoor environments and utilize data from points of interest. In our study, AS is considered a combined outdoor commercial and advertising element, positively associated with SSTR in both clusters. This indirectly validates previous studies' conclusions and further suggests that the management of commercial and public gatherings in these clusters should also consider the risks associated with outdoor environments.

5.3. Limitations and future works

This study encounters certain unavoidable limitations. The SSTR of our analysis was calculated by building with visited/residence COVID-19 confirmed cases, but due to privacy concerns, the data did not break down residents and visitors, and the potential transmission within buildings was not considered. Those limitations may misjudge the building's true SSTR level. The RFA-SHAP interpretable model offers significant explanatory power within the public health sector. However, we identified two weaknesses in our model: the less effective performance in extreme risk scenarios and the suboptimal performance during the 1st and 2nd wave. The reason is that higher model performance demands substantial data volumes for training, but it does not satisfy samples in the 1st and 2nd waves and less extreme risk cases. Given that the GWR shows high fitness in the first two waves, future studies could consider the GWR model for initial modelling and then use the RFA-SHAP model for in-depth analyses. This combined approach may provide more robust and comprehensive results under different risk conditions. Also, model performance under less extreme case conditions may be

beneficial when weighted calculations are applied to the extreme cases.

Additionally, future studies could consider exploring the association between macro-built environment characteristics and regional features, which could provide deeper insights. Attributes such as transport accessibility, the density of high-rise buildings, commercial land concentration, land-use mix, green space density, and building shapes are crucial for understanding SSTR in Hong Kong [12,43]. Integrating SLBE with other BE features could reveal how street-level factors interact with broader urban characteristics to better understand the robustness of SLBE. Additionally, we can explore how SLBE features might have varying impacts depending on the multiple infectious diseases and analyze how these relationships might differ across various phases of a disease, such as the stable, rising, and declining periods.

6. Conclusion

Leveraging advancements in SARS-CoV-2 research, including spatial transmission risk and virus survival strategies, this study utilizes Hong Kong as a case study to introduce an RFA-SHAP model. This model assesses the influence of SLBE features on SSTR within an 800-m walkable service around COVID-19 related buildings during periods of travel restriction. Demonstrating notable efficacy, especially in areas associated with high COVID-19 confirmed cases related to buildings. After fitting the data from the 1st to 5th wave outbreak, the model showed strong predictive ability for HH clusters of COVID-related building clusters, with R^2 test accuracy of 0.86 and RMSE of 0.17. Also, for LL clusters of COVID-related building clusters area, R^2 test accuracy of 0.66 and RMSE of 0.34. The model maintained robustness when fitting individual waves of the outbreak. Key findings and suggestions from this study included.

1) The findings corroborate previous research that SS, SB, and SS are associated with SSTR. In addition, our analysis reveals that in HH cluster areas (e.g., Sheung Wan and Tsim Sha Tsui), street infrastructures such as SSF, SO, and AS play a more substantial role in SARS-CoV-2 transmission, while in LL cluster areas (e.g., Tuen Mun and Tsuen Wan), features like TCF and SSF are more strongly associated with transmission risk. Natural features were more associated with transmission risk during the 1st wave, whereas street infrastructure had a more significant impact during the 3rd and 4th waves.
2) The study suggests a few potential policies recommended for future street management in Hong Kong: (a) For solid waste treatment, especially prioritizing cleaning and managing SSF around street traffic junctions' space; (b) Rational increase of the SO distribution may contribute to reducing SSTR, especially in SS, SSF, AS areas with low and medium distribution density area; and (c) Closing of AS may reduce SSTR by addressing outdoor commercial and public gatherings.

Data availability statement

We acknowledge that all geographical data used in the manuscript were sourced from two open data platforms: Data.Gov.HK (https://data.gov.hk/sc/) and the Hong Kong Common Spatial Data Infrastructure Portal (https://portal.csdi.gov.hk/csdi-webpage/). GSV images were obtained through paid downloads approach via the Google API.

CRediT authorship contribution statement

Chongyang Ren: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Xiaoran Huang:** Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Qingyao Qiao:** Writing – review & editing, Validation, Methodology, Formal analysis. **Marcus White:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

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

Acknowledgments

This research is supported by the National Natural Science Foundation of China [52208039] and the Australian Research Council Linkage Project [LP190100089].

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e38405.

Abbreviation

Artificial Neural Network (ANN)
Artificial Structure (AS)
SARS-CoV-2 Spatial Transmission Risk (SSTR)

Geographically Weighted Regression (GWR)
Google Street View (GSV)
High-High (HH)
High-Low (HL)
Linear Regression (LR)
Light Gradient Boosting Machine (LGBM)
Long short-Term Memory (LSTM)
Low-High (LH)
Low-Low (LL)
Place IDs (PID)
Random Forest (RF)
Random Forest Analysis-SHapley Additive explanations (RFA-SHAP)
R-squared (R^2)
Root Mean Square Error (RMSE)
SHapley Additive explanations (SHAP)
Street Building (SB)
Street Greenness (SG)
Street-level Built Environment (SLBE)
street Obstacles (SO)
Street Public Amenity (SPA)
Street Road Surface (SRU)
Street Sanitation Facility (SSF)
Street Sidewalks (SS)
Street Unpaved Road Surface (SURS)
Support Vector Machine (SVM)
Susceptible-Exposed-Infected-Recovered-Dead (SEIRD)
Traffic Control Facility (TCF)
Tertiary Planning Unit (TPU)
Xgboost (XGBT)

References

[1] A. Zumla, M.S. Niederman, Editorial: the explosive epidemic outbreak of novel coronavirus disease 2019 (COVID-19) and the persistent threat of respiratory tract infectious diseases to global health security, Curr. Opin. Pulm. Med. 26 (3) (2020) 193–196, https://doi.org/10.1097/MCP.0000000000000676.
[2] World Health Organization, COVID-19 weekly epidemiological update. https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19—4-may-2023, 2023.
[3] C.Y.T. Kwok, M.S. Wong, K.L. Chan, et al., Spatial analysis of the impact of urban geometry and socio-demographic characteristics on COVID-19, a study in Hong Kong, Sci. Total Environ. 764 (2021) 144455, https://doi.org/10.1016/j.scitotenv.2020.144455.
[4] N. Pinter-Wollman, A. Jelić, N.M. Wells, The impact of the built environment on health behaviours and disease transmission in social systems, Phil Trans R Soc B 373 (1753) (2018) 20170245, https://doi.org/10.1098/rstb.2017.0245.
[5] K. Samuelsson, S. Barthel, J. Colding, G. Macassa, M. Giusti, Urban nature as a source of resilience during social distancing amidst the coronavirus pandemic, Open Science Framework (2020), https://doi.org/10.31219/osf.io/3wx5a.
[6] Q.C. Nguyen, Y. Huang, A. Kumar, et al., Using 164 million Google street view images to derive built environment predictors of COVID-19 cases, IJERPH 17 (17) (2020) 6359, https://doi.org/10.3390/ijerph17176359.
[7] L. Zhang, X. Han, J. Wu, L. Wang, Mechanisms influencing the factors of urban built environments and coronavirus disease 2019 at macroscopic and microscopic scales: the role of cities, Front. Public Health 11 (2023). https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2023.1137489. (Accessed 15 February 2024).
[8] E.M. Berke, T.D. Koepsell, A.V. Moudon, R.E. Hoskins, E.B. Larson, Association of the built environment with physical activity and obesity in older persons, Am. J. Publ. Health 97 (3) (2007) 486–492, https://doi.org/10.2105/AJPH.2006.085837.
[9] D. Mirabelli, E. Merler, 'Environment' in cancer causation and etiological fraction: limitations and ambiguities, Carcinogenesis 29 (5) (2008) 1083, https://doi.org/10.1093/carcin/bgm305, 1083.
[10] P. Bajardi, C. Poletto, J.J. Ramasco, M. Tizzoni, V. Colizza, A. Vespignani, Human mobility networks, travel restrictions, and the global spread of 2009 H1N1 pandemic, in: M. Perc (Ed.), PLoS One 6 (1) (2011) e16591, https://doi.org/10.1371/journal.pone.0016591.
[11] H. Frumkin, COVID-19, the built environment, and health, Environ. Health Perspect. 129 (7) (2021) 075001, https://doi.org/10.1289/EHP8888.
[12] Z. Kan, M.P. Kwan, M.S. Wong, J. Huang, D. Liu, Identifying the space-time patterns of COVID-19 risk and their associations with different built environment features in Hong Kong, Sci. Total Environ. 772 (2021) 145379, https://doi.org/10.1016/j.scitotenv.2021.145379.
[13] J.M. McLaughlin, F. Khan, S. Pugh, et al., County-level predictors of coronavirus disease 2019 (COVID-19) cases and deaths in the United States: what happened, and where do we go from here? Clin. Infect. Dis. 73 (7) (2021) e1814–e1821, https://doi.org/10.1093/cid/ciaa1729.
[14] S. Hamidi, S. Sabouri, R. Ewing, Does density aggravate the COVID-19 pandemic?: early findings and lessons for planners, J. Am. Plann. Assoc. 86 (4) (2020) 495–509, https://doi.org/10.1080/01944363.2020.1777891.
[15] M. Marquès, J.L. Domingo, Positive association between outdoor air pollution and the incidence and severity of COVID-19. A review of the recent scientific evidences, Environ. Res. 203 (2022) 111930, https://doi.org/10.1016/j.envres.2021.111930.
[16] N. van Doremalen, T. Bushmaker, D.H. Morris, et al., Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1, N. Engl. J. Med. 382 (16) (2020) 1564–1567, https://doi.org/10.1056/NEJMc2004973.
[17] J. Xiong, O. Lipsitz, F. Nasri, et al., Impact of COVID-19 pandemic on mental health in the general population: a systematic review, J. Affect. Disord. 277 (2020) 55–64, https://doi.org/10.1016/j.jad.2020.08.001.

[18] D. Adlakha, J.F. Sallis, Activity-friendly neighbourhoods can benefit non-communicable and infectious diseases, Cities & Health 24 (2020) 1–5, https://doi.org/10.1080/23748834.2020.1783479. Published online July.

[19] N. Naik, J. Philipoom, R. Raskar, C. Hidalgo, Streetscore – predicting the perceived safety of one million streetscapes, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, IEEE, 2014, pp. 793–799, https://doi.org/10.1109/CVPRW.2014.121.

[20] L. Lin, A.V. Moudon, Objective versus subjective measures of the built environment, which are most effective in capturing associations with walking? Health Place 16 (2) (2010) 339–348, https://doi.org/10.1016/j.healthplace.2009.11.002.

[21] Y. Yao, W. Shi, A. Zhang, Z. Liu, S. Luo, Examining the diffusion of coronavirus disease 2019 cases in a metropolis: a space syntax approach, Int. J. Health Geogr. 20 (1) (2021) 17, https://doi.org/10.1186/s12942-021-00270-4.

[22] R.H.M. Pereira, F.R. Borges de Holanda, V.A. Soares de Medeiros, A.P. Barros, The use of space syntax in urban transport analysis: limits and potentials, Discuss. Pap. (2015). https://www.econstor.eu/handle/10419/220276. (Accessed 15 February 2024).

[23] E. Pafka, K. Dovey, G.D. Aschwanden, Limits of space syntax for urban design: axiality, scale and sinuosity, Environ. Plan. B Urban Anal. City Sci. 47 (3) (2020) 508–522, https://doi.org/10.1177/2399808318786512.

[24] A.G. Rundle, M.D.M. Bader, C.A. Richards, K.M. Neckerman, J.O. Teitler, Using Google street view to audit neighborhood environments, Am. J. Prev. Med. 40 (1) (2011) 94–100, https://doi.org/10.1016/j.amepre.2010.09.034.

[25] I. Seiferling, N. Naik, C. Ratti, R. Proulx, Green streets – Quantifying and mapping urban trees with street-level imagery and computer vision, Landsc. Urban Plann. 165 (2017) 93–101, https://doi.org/10.1016/j.landurbplan.2017.05.010.

[26] P. Griew, M. Hillsdon, C. Foster, E. Coombes, A. Jones, P. Wilkinson, Developing and testing a street audit tool using Google Street View to measure environmental supportiveness for physical activity, Int. J. Behav. Nutr. Phys. Activ. 10 (1) (2013) 103, https://doi.org/10.1186/1479-5868-10-103.

[27] L. Phan, W. Yu, J.M. Keralis, et al., Google street view derived built environment indicators and associations with state-level obesity, physical activity, and chronic disease mortality in the United States, IJERPH 17 (10) (2020) 3659, https://doi.org/10.3390/ijerph17103659.

[28] V. Silva, A. Grande, C. Rech, M. Peccin, Geoprocessing via Google maps for assessing obesogenic built environments related to physical activity and chronic noncommunicable diseases: validity and reliability, Journal of Healthcare Engineering 6 (1) (2015) 41–54, https://doi.org/10.1260/2040-2295.6.1.41.

[29] Q.C. Nguyen, J.M. Keralis, P. Dwivedi, et al., Leveraging 31 million Google street view images to characterize built environments and examine county health outcomes, Publ. Health Rep. 136 (2) (2021) 201–211, https://doi.org/10.1177/0033354920968799.

[30] Q.C. Nguyen, T. Belnap, P. Dwivedi, et al., Google street view images as predictors of patient health outcomes, 2017–2019, BDCC 6 (1) (2022) 15, https://doi.org/10.3390/bdcc6010015.

[31] Hassaan MA, Abdelwahab RG, Elbarky TA, Ghazy RM. GIS-based analysis framework to identify the determinants of COVID-19 incidence and fatality in Africa. Community Health. .

[32] R.C. Urban, L.Y.K. Nakada, GIS-based spatial modelling of COVID-19 death incidence in São Paulo, Brazil, Environ. Urbanization 33 (1) (2021) 229–238, https://doi.org/10.1177/0956247820963962.

[33] A. Maiti, Q. Zhang, S. Sannigrahi, et al., Exploring spatiotemporal effects of the driving factors on COVID-19 incidences in the contiguous United States, Sustain. Cities Soc. 68 (2021) 102784, https://doi.org/10.1016/j.scs.2021.102784.

[34] S. Arifin, E.T. Herdiani, Modeling of COVID-19 Cases in Indonesia with the Method of Geographically Weighted Regression. J 19 (2) (2023) 342–350, https://doi.org/10.20956/j.v19i2.23481.

[35] F. Liu, J. Wang, J. Liu, et al., Predicting and analyzing the COVID-19 epidemic in China: based on SEIRD, LSTM and GWR models, PLoS One 15 (8) (2020) e0238280, https://doi.org/10.1371/journal.pone.0238280.

[36] S. Ma, S. Li, J. Zhang, Diverse and nonlinear influences of built environment factors on COVID-19 spread across townships in China at its initial stage, Sci. Rep. 11 (1) (2021) 1–13, https://doi.org/10.1038/s41598-021-91849-1.

[37] Y. Han, J. Huang, R. Li, et al., Impact analysis of environmental and social factors on early-stage COVID-19 transmission in China by machine learning, Environ. Res. 208 (2022) 112761, https://doi.org/10.1016/j.envres.2022.112761.

[38] Transport department - the annual traffic census 2019. https://www.td.gov.hk/en/publications_and_press_releases/publications/free_publications/the_annual_traffic_census_2019/index.html, October 6, 2022. (Accessed 17 February 2024).

[39] B.J. Cowling, S.T. Ali, T.W.Y. Ng, et al., Impact assessment of non-pharmaceutical interventions against coronavirus disease 2019 and influenza in Hong Kong: an observational study, Lancet Public Health 5 (5) (2020) e279–e288, https://doi.org/10.1016/S2468-2667(20)30090-6.

[40] F.K. Cheng, Community health awareness model: discussion revisited on how Hong Kong people are combating the COVID-19 pandemic, Moroccan Journal of Public Heath 2 (1) (2021) 9–21.

[41] N. Zhang, W. Jia, H. Lei, et al., Effects of human behavior changes during the coronavirus disease 2019 (COVID-19) pandemic on influenza spread in Hong Kong, Clin. Infect. Dis. 73 (5) (2021) e1142–e1150, https://doi.org/10.1093/cid/ciaa1818.

[42] N. Zhang, W. Jia, P. Wang, et al., Changes in local travel behaviour before and during the COVID-19 pandemic in Hong Kong, Cities 112 (2021) 103139, https://doi.org/10.1016/j.cities.2021.103139.

[43] J. Huang, M.P. Kwan, Z. Kan, M. Wong, C. Kwok, X. Yu, Investigating the relationship between the built environment and relative risk of COVID-19 in Hong Kong, IJGI 9 (11) (2020) 624, https://doi.org/10.3390/ijgi9110624.

[44] M. Santos-Vega, M.J. Bouma, V. Kohli, M. Pascual, Population density, climate variables and poverty synergistically structure spatial risk in urban malaria in India, in: M. Maheu-Giroux (Ed.), PLoS Neglected Trop. Dis. 10 (12) (2016) e0005155, https://doi.org/10.1371/journal.pntd.0005155.

[45] L. Chen, Y. Lu, Q. Sheng, Y. Ye, R. Wang, Y. Liu, Estimating pedestrian volume using Street View images: a large-scale validation test, Comput. Environ. Urban Syst. 81 (2020) 101481, https://doi.org/10.1016/j.compenvurbsys.2020.101481.

[46] X. Ma, C. Ma, C. Wu, et al., Measuring human perceptions of streetscapes to better inform urban renewal: a perspective of scene semantic parsing, Cities 110 (2021) 103086, https://doi.org/10.1016/j.cities.2020.103086.

[47] L. Yang, J. Liu, Y. Liang, Y. Lu, H. Yang, Spatially varying effects of street greenery on walking time of older adults, IJGI 10 (9) (2021) 596, https://doi.org/10.3390/ijgi10090596.

[48] X. Fu, T. Jia, X. Zhang, S. Li, Y. Zhang, Do street-level scene perceptions affect housing prices in Chinese megacities? An analysis using open access datasets and deep learning, in: S. Fu (Ed.), PLoS One 14 (5) (2019) e0217505, https://doi.org/10.1371/journal.pone.0217505.

[49] H. Tahir, M. Shahbaz Khan, M. Owais Tariq, Performance analysis and comparison of faster R-CNN, Mask R-CNN and ResNet50 for the detection and counting of vehicles, in: 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), IEEE, 2021, pp. 587–594, https://doi.org/10.1109/ICCCIS51004.2021.9397079.

[50] L. Zhang, J. Wu, Y. Fan, H. Gao, Y. Shao, An efficient building extraction method from high spatial resolution remote sensing images based on improved Mask R-CNN, Sensors 20 (5) (2020) 1465, https://doi.org/10.3390/s20051465.

[51] S. Nagata, T. Nakaya, T. Hanibuchi, S. Amagasa, H. Kikuchi, S. Inoue, Objective scoring of streetscape walkability related to leisure walking: statistical modeling approach with semantic segmentation of Google Street View images, Health Place 66 (2020) 102428, https://doi.org/10.1016/j.healthplace.2020.102428.

[52] Y. Lu, The association of urban greenness and walking behavior: using Google street view and deep learning techniques to estimate residents' exposure to urban greenness, IJERPH 15 (8) (2018) 1576, https://doi.org/10.3390/ijerph15081576.

[53] E. Boakye-Dankwa, A. Nathan, A. Barnett, et al., Walking behaviour and patterns of perceived access to neighbourhood destinations in older adults from a low-density (Brisbane, Australia) and an ultra-dense city (Hong Kong, China), Cities 84 (2019) 23–33, https://doi.org/10.1016/j.cities.2018.07.002.

[54] Y. Tian, C. Jim, Y. Liu, Using a spatial interaction model to assess the accessibility of district parks in Hong Kong, Sustainability 9 (11) (2017) 1924, https://doi.org/10.3390/su9111924.

[55] J. Miles, Tolerance and variance inflation factor, in: R.S. Kenett, N.T. Longford, W.W. Piegorsch, F. Ruggeri (Eds.), Wiley StatsRef: Statistics Reference Online, first ed., Wiley, 2014. https://onlinelibrary.wiley.com/doi/10.1002/9781118445112.stat06593. (Accessed 1 August 2024).

[56] A. Garg, K. Tai, Comparison of statistical and machine learning methods in modelling of data with multicollinearity, IJMIC 18 (4) (2013) 295, https://doi.org/10.1504/IJMIC.2013.053535.

[57] J.I. Daoud, Multicollinearity and regression analysis, J. Phys.: Conf Ser. 949 (2017) 012009, https://doi.org/10.1088/1742-6596/949/1/012009.
[58] Xiong Y. Spatial Statistics and in Uencing Factors of the Novel Coronavirus Pneumonia 2019 Epidemic in Hubei Province, China. .
[59] N. Aral, H. Bakir, Spatiotemporal analysis of covid-19 in Turkey, Sustain. Cities Soc. 76 (2022) 103421, https://doi.org/10.1016/j.scs.2021.103421.
[60] Z. Xie, Y. Qin, Y. Li, W. Shen, Z. Zheng, S. Liu, Spatial and temporal differentiation of COVID-19 epidemic spread in mainland China and its influencing factors, Sci. Total Environ. 744 (2020) 140929, https://doi.org/10.1016/j.scitotenv.2020.140929.
[61] J.K. Ord, A. Getis, Local spatial autocorrelation statistics: distributional issues and an application, Geogr. Anal. 27 (4) (1995) 286–306, https://doi.org/10.1111/j.1538-4632.1995.tb00912.x.
[62] X. Xu, W. Qiu, W. Li, et al., Associations between street-view perceptions and housing prices: subjective vs. Objective measures using computer vision and machine learning techniques, Rem. Sens. 14 (4) (2022) 891, https://doi.org/10.3390/rs14040891.
[63] Bergstra J, Bergstra J, Bengio Y, Bengio Y. Random Search for Hyper-Parameter Optimization. .
[64] C. Willmott, K. Matsuura, Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, Clim. Res. 30 (2005) 79–82, https://doi.org/10.3354/cr030079.
[65] J. Lupón, H.K. Gaggin, M. de Antonio, et al., Biomarker-assist score for reverse remodeling prediction in heart failure: the ST2-R2 score, Int. J. Cardiol. 184 (2015) 337–343, https://doi.org/10.1016/j.ijcard.2015.02.019.
[66] Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. .
[67] Rubinstein RY. SIMULATION AND THE MONTE CARL0 METHOD. .
[68] S. Bouazizi, H. Ltifi, Enhancing accuracy and interpretability in EEG-based medical decision making using an explainable ensemble learning framework application for stroke prediction, Decis. Support Syst. 178 (2024) 114126, https://doi.org/10.1016/j.dss.2023.114126.
[69] Y. Yao, H. Yin, C. Xu, et al., Assessing myocardial infarction severity from the urban environment perspective in Wuhan, China, J. Environ. Manag. 317 (2022) 115438, https://doi.org/10.1016/j.jenvman.2022.115438.
[70] C. Brunsdon, S. Fotheringham, M. Charlton, Geographically weighted regression, J. Roy. Stat. Soc. D 47 (3) (1998) 431–443, https://doi.org/10.1111/1467-9884.00145.
[71] A. Saefuddin, N.A. Setiabudi, A. Fitrianto, On Comparison between Logistic Regression and Geographically Weight (2012). Published online.
[72] L. Giusti, A review of waste management practices and their impact on human health, Waste Manag. 29 (8) (2009) 2227–2239, https://doi.org/10.1016/j.wasman.2009.03.028.
[73] G. Rachiotis, D. Papagiannis, D. Markas, E. Thanasias, G. Dounias, C. Hadjichristodoulou, Hepatitis B virus infection and waste collection: prevalence, risk factors, and infection pathway, Am. J. Ind. Med. 55 (7) (2012) 650–655, https://doi.org/10.1002/ajim.22057.
[74] S. Needhidasan, M. Samuel, R. Chidambaram, Electronic waste – an emerging threat to the environment of urban India, J Environ Health Sci Engineer 12 (1) (2014) 36, https://doi.org/10.1186/2052-336X-12-36.
[75] F. Di Maria, E. Beccaloni, L. Bonadonna, et al., Minimization of spreading of SARS-CoV-2 via household waste produced by subjects affected by COVID-19 or in quarantine, Sci. Total Environ. 743 (2020) 140803, https://doi.org/10.1016/j.scitotenv.2020.140803.
[76] A.E. Aiello, R.M. Coulborn, V. Perez, E.L. Larson, Effect of hand hygiene on infectious disease risk in the community setting: a meta-analysis, Am. J. Publ. Health 98 (8) (2008) 1372–1381, https://doi.org/10.2105/AJPH.2007.124610.
[77] D.K. Chu, E.A. Akl, S. Duda, et al., Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis, Lancet 395 (10242) (2020) 1973–1987, https://doi.org/10.1016/S0140-6736(20)31142-9.
[78] S. Jamaludin, N.A. Azmir, A.F. Mohamad Ayob, N. Zainal, COVID-19 exit strategy: transitioning towards a new normal, Annals of Medicine and Surgery 59 (2020) 165–170, https://doi.org/10.1016/j.amsu.2020.09.046.
[79] L. Dietz, P.F. Horve, D.A. Coil, M. Fretz, J.A. Eisen, K. Van Den Wymelenberg, Novel coronavirus (COVID-19) pandemic: built environment considerations to reduce transmission, mSystems 5 (2) (2019) e00245, https://doi.org/10.1128/mSystems.00245-20, 2020.