



# The Analytical Workflow for Shifts in Post-COVID Living Preferences on Neighborhood Qualities

Zhiyi Dou<sup>1</sup>, Waishan Qiu<sup>2</sup>, Wenjing Li<sup>3</sup>, and Dan Luo<sup>4</sup>(✉)

<sup>1</sup> School of Geographic Information Science, University of Queensland, St. Lucia, QLD, Australia

<sup>2</sup> Department of City and Regional Planning, Cornell University, Ithaca, NY, USA

<sup>3</sup> Center for Spatial Information Science, The University of Tokyo, Tokyo, Japan

<sup>4</sup> School of Architecture, University of Queensland, St. Lucia, QLD, Australia  
d.luo@uq.edu.au

**Abstract.** potential self-organization of work-life patterns and social profiles in the designated neighbourhood.

To evaluate the subjective perception of the urban residence, the study started with a comparative survey by asking residence to compare two randomly selected urban contexts in a data base of 398 contexts sampled across Hong Kong and state their living preference under the presumption of following scenarios: 1. working from home; 2. working in city centre offices. Core information influencing the spatial equilibrium are provided in the comparable urban context such as street views, housing price, housing space, travel time to city centre, adjacency to public transport and amenities, etc. Each context is given a preference score calculated with Microsoft TrueSkill Bayesian ranking algorithm based on the comparison survey of two scenarios.

The 398 contexts are further analysed via GIS and image processing, to be deconstructed into numerical values describing main features for each of the context that influence urban design strategies such as composition of spatial features, amenity allocation, adjacency to city centre and public transportations. Machine learning models are trained with the numerical values of urban features as input and two preference scores for the two working scenarios as the output. The correlation heat maps are used to identify main urban features and its p-value that influence residence's preference under two working scenarios in post-COVID era. The same model could also be applied to inform the direction of urban design strategies to construct a sustainable community for each type of working population and validate the design strategies via predicting its competitiveness in attracting residence and developing target industries.

**Keywords:** Urban data analysis · Sustainable communities · Post-COVID · Data driven urban design · Machine learning

## 1 Introduction

The post-COVID era has influenced the way cities to operate. Workplace practices and modes of collaboration have subtly changed the preferences of the working population for urban existence. Modern cities are designed and developed to provide humans with a convenient and healthy environment. Cities are densely populated with commercial and industrial development, leading to vulnerability to epidemic outbreaks [5]. More companies are allowing their employees to work remotely permanently instead of in physical offices to cope with the effects of this global epidemic [2]. This change in work patterns has led to a mismatch between the spatial form of the city and the matching population in the post-COVID period. Residents will need to consider their work and lifestyle during Lockdown. When some work patterns shift to remote working, this has implications for mobility, transportation, and housing prices. The working population will face two scenarios. First, the convenience premium enjoyed by residents of neighbouring cities with short commutes and easy commutes will disappear for people working from home. Some workers may tolerate longer commutes from further afield if they only need to commute to the city center once or twice a week [6]. Fewer people coming to offices can dampen demand for office space, especially in expensive inner-city areas. Under the circumstance of ensuring the public health and safety of the city, the spatial distribution of urban resources will have two extremes of scarcity and excess. Therefore, urban design and construction will need to reconsider people's needs. Considering the interweaving of urban context at multiple scales and complex decision-making process based on subjective preference, the current methods in constructing modelled predicting the performance of urban development face the following key challenges:

1. The difficulties to evaluate the impact of urban features across multiple scales, such as urban planning qualities and local spatial experience.
2. Challenge to incorporating citizen's subjective opinions into the evaluation model.
3. How to expand a model trained with inputs from current context to predict the performance of future novel scenarios.
4. People's spontaneous activity may deviate from their stated preference from survey.

Based on survey and case study of Hong Kong, this research aims to address the key challenge by developing a comprehensive workflow with the following features:

1. Urban features from both planning data and urban street views are incorporated into the evaluation with learnable weight.
2. Intuitive subjective rating from users is applied to generate preference score.
3. A portion of the data for the sampled points are manipulated to include scenarios for future development via statistical analysis of data distribution in the design space.
4. Data mined from spontaneous activities such as real-estate rental and social media is used to validate the accuracy of the stated preference from the survey, and evaluate how faithful they are in reflecting people choose realistically.

## 2 Literature Review

As an essential part of urban public space, the street is not only the primary carrier of transportation, but also an important space for personal daily activities, including work and communication [7]. Street environment is a major factor influencing people's physical activity, the choice to move and willingness to pay [3]. Previous assessments of the street environment were mainly derived from two measures. The first was objective based data of city to measure street quality, such as building height, park area, number of surrounding transportation facilities, etc. However, a single objective measurement cannot represent the overall feeling of human residents. Another is a subjective measurement, such as interviews or surveys [8]. However, due to individual differences, the consistency and reliability of operations will influence the uncertainty of the results. Therefore, methods that combine subjective perception scores and field surveys to obtain objective measurement data have emerged. For example, Qiu et al. applied computer vision and machine learning to subjectively measure four perceptual qualities collected from SVI samples using Shanghai [3].

In the post-COVID period, changes in cities are mainly changes in the spatial distance of people, affecting living choices, travel, and commuting patterns. The link between COVID-19 prevalence and urban design features has sparked much debate in some studies. However, the existing literature does not detail which street environmental factors influence residents' preferences during the post-COVID period [6]. Therefore, our work aims to study the impact of work patterns on their preferences by targeting the working population of a company in a city center. Based on extracting the pixel ratio of objective street view elements using SVI data, adding urban spatial information to enrich urban perception to achieve subjective measurement. The real estate data is then used to validate the score. The results will show two work scenarios affecting residents' preferences in the post-COVID era. The same model will also inform post-pandemic urban design strategies.

## 3 Data and Methodology

This study attempts to take the current Hong Kong city as an example in response to the above issues. The street image of the study area is segmented to obtain the urban feature values, and the city's physical environment is perceived through the street view image. After adding objective urban spatial data, enriched investigators' subjective choice of residence preferences for workplaces in a randomly sampled Street View database. We use machine learning to predict future urban development design strategies in the post-COVID era. This research will construct a sustainable urban environment and future development direction for the working population by evaluating this model.

A key characteristic for the workflow of this methodology is that both urban context from planning scale such as the POI distribution, abundance of public transportation, parks and distance to city centre, as well as local urban quality such as street experience is taken into consideration altogether. Opinions of citizen in incorporated via a subjective rating process. The rating is based on manipulated data with improved distribution in design space to incorporate future scenarios of development. Lastly, data revealing

behaviour and choice in real world, such as real-estate rental data is applied to validate the accuracy of the preference stated by the survey takers and evaluate the extend for them to forecast people’s behaviours and opinions on current and future development (Fig. 1).

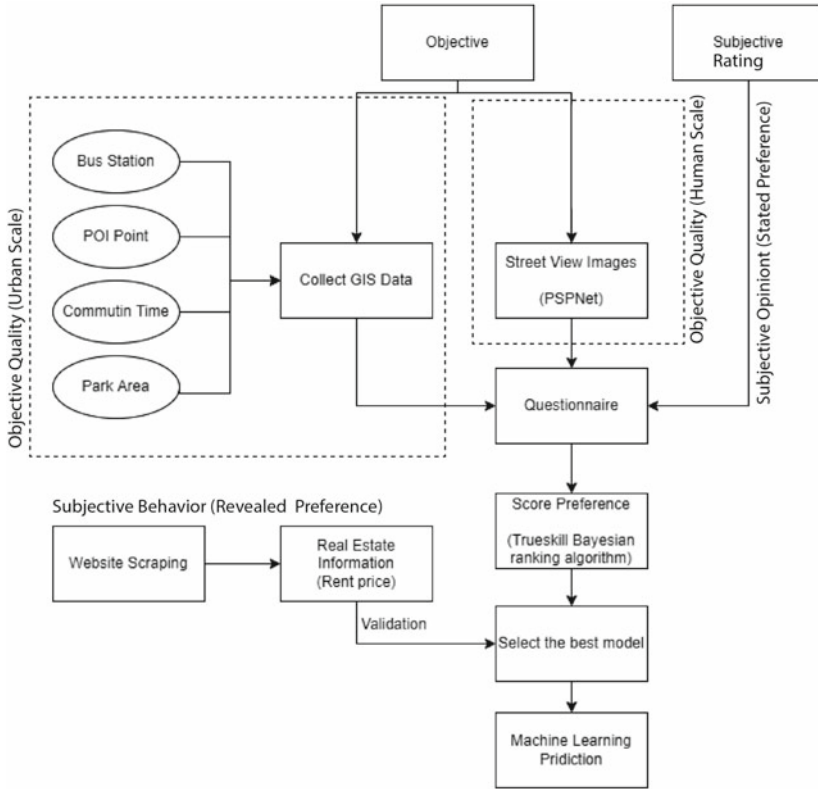


Fig. 1. Analytical framework

### 3.1 Study Area

The Hong Kong Special Administrative Region, located at the southern tip of mainland China, was selected as the study area. The whole territory consists of Hong Kong Island, Kowloon, and the New Territories. It is a highly prosperous free port and cosmopolitan city. With a total population of more than 7 million and an area of about 1,100 square kilometers, it is also the third-largest financial center in the world. Hong Kong has a unique geographical space environment. Due to its mountainous terrain, dense urban population and numerous high-rise buildings, it only accounts for about 20% of the general land area [9]. The analysis of the street quality in Hong Kong will provide inspiration for future urban renewal.

The objective spatial data of the study area include: (1) The distribution of Hong Kong bus stations, POI, and park areas from HONG KONG GEODATA STORE (<https://geodata.gov.hk/gis/>), (2) a shapefile of road networks from Open Street Map (<https://www.openstreetmap.org/>), (3) SVI collected from Google Maps Street View API, (4) property information from Centaline (<https://hk.centanet.com/info/index>).

### 3.2 Data Collection Processing

The road network in the form of linear vector data is extracted from the open source of Open Street Map, then converted into shapefile by QGIS to obtain all road network data in Hong Kong. We are creating 1,000 points in the road network data for each district in Hong Kong with an interval of 50 m and getting 18,000 points. Ensure that our training images cover most urban area types and improve efficiency. We randomly selected 398 study sites. Then according to the geographic coordinates of the research point, download the corresponding SVI from the Google Street View static API (Fig. 2).

Objective factors are significant criteria to measure street quality. We selected “number of bus stops,” “number of POIs,” “size of park area,” and “commute time” to present the perceived streetscape quality. Processing in the geographic information software ArcGIS Pro. A buffer zone of 1.5 km<sup>2</sup> is established with randomly selected research sites as the center. The number of bus stops in the buffer zone of each study site, the number of POIs, and the park area covered are counted. The commute time was then estimated based on the distance from the study site to the city center. The number of bus stops represents the degree of crucial urban infrastructure and is one of the criteria for people to judge the accessibility of a place of residence [10]. The number of POIs represents the urban spatial structure and provides scientific support for guiding regional spatial optimization and regulation. POI data records the spatial location information of various social and economic sectors, which can more clearly identify the types provided by urban functional areas [11]. The area of parks represents an essential indicator of the overall environmental level of the city and the quality of life of residents [12]. Commuting time represents the travel behavior of residents who live and work apart and is an essential factor affecting the residential experience of the working population.

After collecting context data based on existing scenario, a noise and variance is applied to data set, so that the context parameters are distributed across the design parameter space, enable the dataset to encompass future scenarios in addition to existing context (Fig. 3).

### 3.3 Physical Feature at Personal Scale

People’s preference scores for Street View were collected through an online questionnaire platform developed (Fig. 4). The online survey system asked participants to click on one image in a pair of SVIs to answer the assessment questions. Street view quality is perceived through human vision, and provides urban spatial quality data to deepen judgment. If the company is in the city center, the participants will assume that they are in two different environments: choose the living area based on working from home or the office. The 398 SVIs mentioned above randomly sampled from the Hong Kong study area were used in the survey. A total of 1900 valid contributions were collected, and

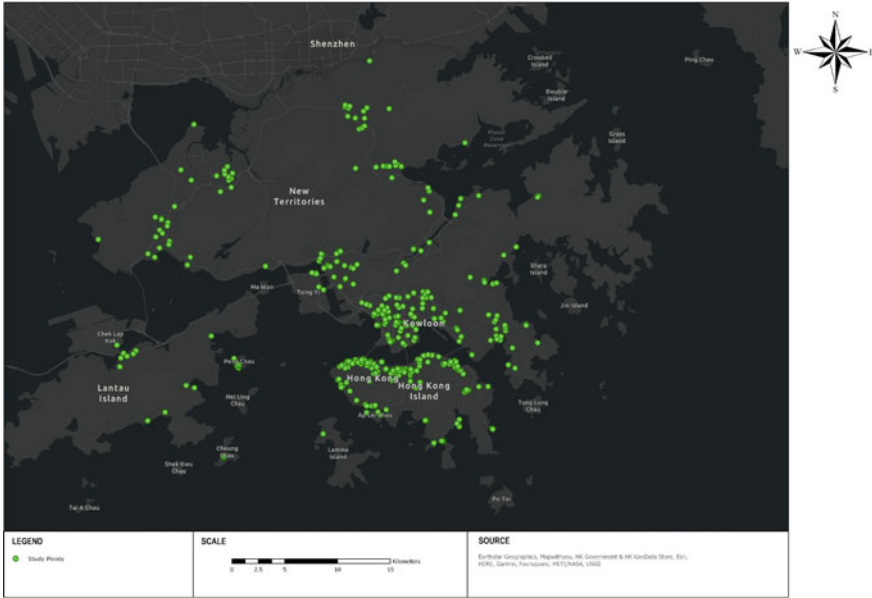


Fig. 2. 398 sampled study points



Fig. 3. Study points scatter plot

the participants were architecture students at the University of Queensland. Averaged to ensure that each image participates in more than 5 comparisons.

Part of the physical features in the street view is one of the essential factors for human perception of the quality of the street view [3]. The pixel ratio of a single feature in SVI is obtained through the Pyramid Scene Parsing Network (PSPNet). It is a pixel-level object recognition and classification algorithm with over 93.4% pixel-level accuracy. More than 30 Street View elements were detected in the study area. Since the uncertain number changes of pedestrians, cars and some facilities and the size of the covered image area have little significance to the pixel ratio, only eight meaningful street elements are selected as the primary reference standards.

To faithfully translate citizens opinions into statistical values, Microsoft True Skill ranking is applied to generate scores based on multiple comparison outcomes between randomly sampled urban scenarios. In order to generate the relative scores that reflects

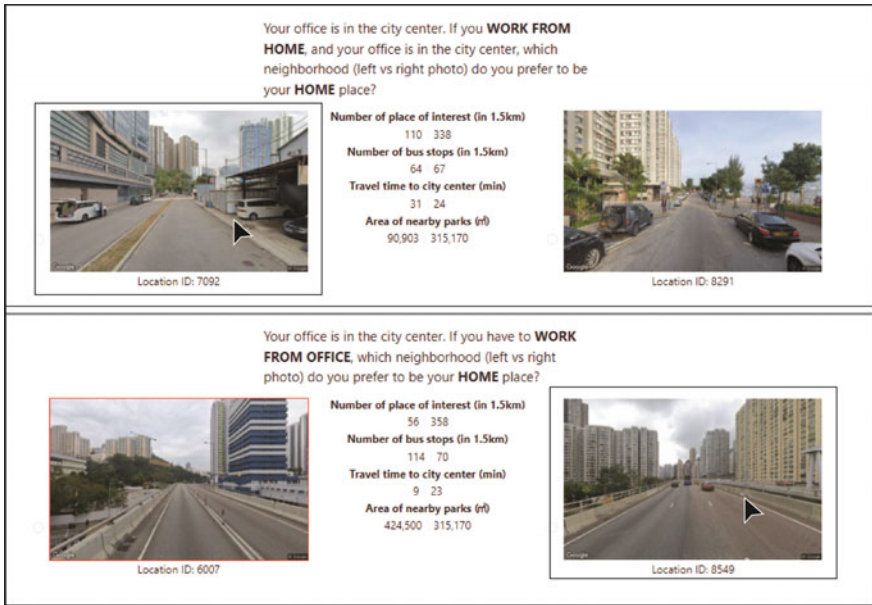


Fig. 4. Questionnaire/Survey for Revealing Subjective Preference

the ranking of preference for individual scenarios, each scenario composed of a set of urban information and a street view photo should be compared with at least three other random scenarios. The benefit of using a dynamic comparison survey instead of a static rating survey is to minimize human base bias in rating, such as some people tends to give higher rating comparing to others (Fig. 5).

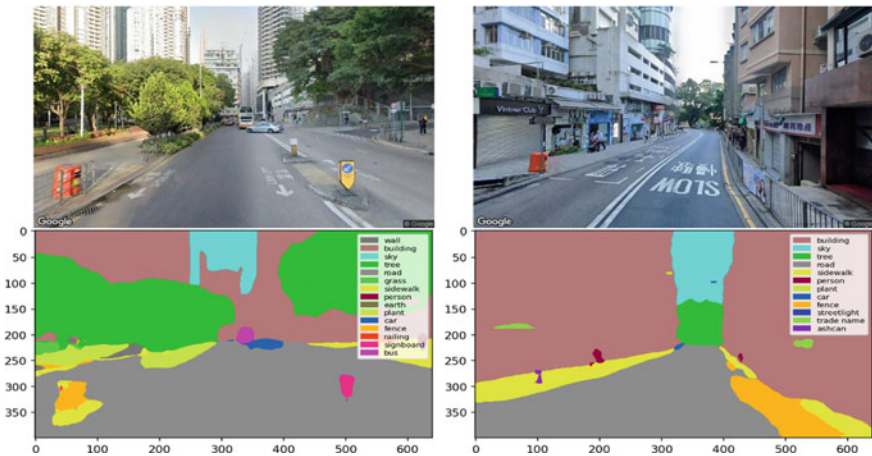


Fig. 5. PSPNet semantic segmentation results with raw input (study data 13 and 510)

### 3.4 Model Training and Validation

The target output variables of this study are two preference scores. The above-mentioned street view elements obtained from SVI are fed into the machine learning model. Refer to Qiu et al. for the Pudong Area related research using 8 ML algorithms [3]. They are Linear Regression, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Random Forest (RF), Decision Tree (DT), Voting Selection (VS), Gradient Boosting (GB) and Adaptive Boosting (ADAB), to Predict two preference scores. However, traditional machine learning method primarily working with higher lever features that require pre-processing of the street scene itself, such as extracted image segmentation value representing the percentage of urban element. This pre-processing of the image, though able to extract readable urban feature qualities, is in-efficient to translate the more abstract qualities such as style and atmosphere. Thus, in addition to applying traditional machine learning algorithm, the research has progressed toward applying Convolutional Neural Networks as the foundation of constructing the perdition model.

Set the Mean Square Error (MSE) as the loss function to evaluate the performance of different ML and Neural Network algorithm. It shows advantages in evaluating average models. The lower MSE value implies the higher accuracy of a prediction model. In addition, R-Square is a key parameter determine the proportion of the variance for a dependent variable that’s explained by an independent variable or variables in a regression model.

Multiple models with traditional machine learning is applied including K-Nearest Neighbors (KNN), Support Vector Machine (SVM, Random Forest etc. as well as a customized Neural network with Convolutional layers (CNN). The better performing traditional machine learning method include KNN and SVM achieved RMSE of 0.21 and R2 of 0.24, while CNN is able to significantly improve the performance toward RMSE of 0.153 and R2 of 0.476. The performance of CNN has proven to be reaching the level of applicability in for predicting the human preference of living condition under the work from office scenario (Fig. 6).

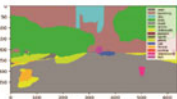

Traditional Machine Learning Methods		Neural Networks (CNN)																																																																										
<p>Input1: Street Scene Features</p> 	<p>Input2: Urban Context</p> <ul style="list-style-type: none"> <li>POI count (1.5km radius)</li> <li>Bus station Count (1.5km radius)</li> <li>Park area (1.5km radius)</li> <li>Travel time to city center</li> </ul>	<p>Input1: Street Scene Image</p> 	<p>Input2: Urban Context</p> <ul style="list-style-type: none"> <li>POI count (1.5km radius)</li> <li>Bus station Count (1.5km radius)</li> <li>Park area (1.5km radius)</li> <li>Travel time to city center</li> </ul>																																																																									
Output: Work in office preference score		Output: Work in office preference score																																																																										
<p>Model Type and Performance</p> <table border="1"> <thead> <tr> <th>Model Type</th> <th>KNN</th> <th>SVM</th> <th>Random Forest</th> <th>Decision Tree</th> <th>SLS</th> <th>Selection</th> <th>Voting Selection</th> <th>Gradient Boosting</th> <th>ADA Boost</th> <th>Bagging Regression</th> </tr> </thead> <tbody> <tr> <td>R2</td> <td>0.24</td> <td>0.24</td> <td>0.26</td> <td>0.16</td> <td>0.16</td> <td>0.18</td> <td>0.18</td> <td>0.2</td> <td>0.18</td> <td>0.21</td> </tr> <tr> <td>RMSE</td> <td>0.21</td> <td>0.21</td> <td>0.21</td> <td>0.22</td> <td>0.22</td> <td>0.22</td> <td>0.22</td> <td>0.21</td> <td>0.22</td> <td>0.21</td> </tr> </tbody> </table>		Model Type	KNN	SVM	Random Forest	Decision Tree	SLS	Selection	Voting Selection	Gradient Boosting	ADA Boost	Bagging Regression	R2	0.24	0.24	0.26	0.16	0.16	0.18	0.18	0.2	0.18	0.21	RMSE	0.21	0.21	0.21	0.22	0.22	0.22	0.22	0.21	0.22	0.21	<p>Model Structure</p> <table border="1"> <thead> <tr> <th>Layer (Type)</th> <th>Output Shape</th> <th>Param #</th> <th>Connected to</th> </tr> </thead> <tbody> <tr> <td>input_25 (InputLayer)</td> <td>(None, 80, 128, 3)</td> <td>0</td> <td>[]</td> </tr> <tr> <td>input_26 (InputLayer)</td> <td>(None, 512)</td> <td>0</td> <td>[]</td> </tr> <tr> <td>sequential_12 (Sequential)</td> <td>(None, 51)</td> <td>16805</td> <td>input_25(input)</td> </tr> <tr> <td>concatenate_12 (Concatenate)</td> <td>(None, 101)</td> <td>0</td> <td>input_26(input); sequential_12(output)</td> </tr> <tr> <td>dense_72 (Dense)</td> <td>(None, 16)</td> <td>176</td> <td>concatenate_12(output)</td> </tr> <tr> <td>dense_73 (Dense)</td> <td>(None, 8)</td> <td>136</td> <td>dense_72(output)</td> </tr> <tr> <td>dense_74 (Dense)</td> <td>(None, 8)</td> <td>72</td> <td>dense_73(output)</td> </tr> <tr> <td>dense_75 (Dense)</td> <td>(None, 4)</td> <td>36</td> <td>dense_74(output)</td> </tr> <tr> <td>dense_76 (Dense)</td> <td>(None, 1)</td> <td>5</td> <td>dense_75(output)</td> </tr> </tbody> </table> <p>Total params: 17,230                  Trainable params: 17,230                  Non-trainable params: 0</p> <p>Model Performance                  R2 = 0.476, RMSE = 0.153</p>		Layer (Type)	Output Shape	Param #	Connected to	input_25 (InputLayer)	(None, 80, 128, 3)	0	[]	input_26 (InputLayer)	(None, 512)	0	[]	sequential_12 (Sequential)	(None, 51)	16805	input_25(input)	concatenate_12 (Concatenate)	(None, 101)	0	input_26(input); sequential_12(output)	dense_72 (Dense)	(None, 16)	176	concatenate_12(output)	dense_73 (Dense)	(None, 8)	136	dense_72(output)	dense_74 (Dense)	(None, 8)	72	dense_73(output)	dense_75 (Dense)	(None, 4)	36	dense_74(output)	dense_76 (Dense)	(None, 1)	5	dense_75(output)
Model Type	KNN	SVM	Random Forest	Decision Tree	SLS	Selection	Voting Selection	Gradient Boosting	ADA Boost	Bagging Regression																																																																		
R2	0.24	0.24	0.26	0.16	0.16	0.18	0.18	0.2	0.18	0.21																																																																		
RMSE	0.21	0.21	0.21	0.22	0.22	0.22	0.22	0.21	0.22	0.21																																																																		
Layer (Type)	Output Shape	Param #	Connected to																																																																									
input_25 (InputLayer)	(None, 80, 128, 3)	0	[]																																																																									
input_26 (InputLayer)	(None, 512)	0	[]																																																																									
sequential_12 (Sequential)	(None, 51)	16805	input_25(input)																																																																									
concatenate_12 (Concatenate)	(None, 101)	0	input_26(input); sequential_12(output)																																																																									
dense_72 (Dense)	(None, 16)	176	concatenate_12(output)																																																																									
dense_73 (Dense)	(None, 8)	136	dense_72(output)																																																																									
dense_74 (Dense)	(None, 8)	72	dense_73(output)																																																																									
dense_75 (Dense)	(None, 4)	36	dense_74(output)																																																																									
dense_76 (Dense)	(None, 1)	5	dense_75(output)																																																																									

Fig. 6. Comparison between the performance of traditional machine learning algorithm and CNN



## 4 Conclusions and Feature Work

This study applies the trade-offs of urban residents' perception of urban spatial quality and urban public facilities in the prediction environment of ML models. The analysis was performed using QGIS, ArcGIS Pro and Python. The study areas were 398 study areas randomly selected from Hong Kong. From a theoretical point of view, we use the SVI segmented pixel ratio dataset to replace the subjective measurement of human eyes perception on the one hand, and then add objectively measured urban spatial quality data. The results will be more comprehensive and nuanced to reflect the working population's residency preferences.

The advantage of this evaluation model is that it combines subjective measurements of objective factors, which can more comprehensively predict and guide the design of the street environment in the post-COVID period. Since the dataset fits the pedestrian's perspective, the cost is meagre, and it does not require highly complex technology and can be used only in the area where the SVI dataset is available. According to the target population, such as the working population, relevant spatial elements (bus stops, parks, POI, commute time) are added to improve the accuracy of the prediction results.

Kindly note, this research is proof-of-concept research for the methodology and workflow, with preference data collected from design students. The preference data does not reflect the opinion of Hong Kong citizen and will not be published. In addition, The accuracy of PSPNet parsing SVI images is not 100%, though using NN and inputting realistic urban photo would improve this error. Then the environment of SVI can only be used in common scenes, and lack of analysis of specific environments, such as some private streets and interior scenes, cannot obtain images. For the 398 SVIs with a limited number of participants and randomly selected, the accuracy and reliability of the prediction results still need to be improved. These will also be the main directions of future research work, providing more complete and precise guidance for the sustainable development of urban construction. However, with the validation of the workflow, the current model has validated its performance statistically, thus proving this methodology can be applied toward multiple cities given a systematic survey with sampled targeted user group and additional urban samplings.

## References

1. Gollin D, Kirchberger M, Lagakos D (2017) In search of a spatial equilibrium in the developing world. *Nat Bureau Econ Res*
2. Miranda AS, Fan Z, Duarte F, Ratti C (2021) Desirable streets: using deviations in pedestrian trajectories to measure the value of the built environment. *Comput Environ Urban Syst*
3. Qiu W, Li W, Liu X et al (2021) Subjectively measured streetscape perceptions to inform urban design strategies for Shanghai. *ISPRS Int J Geo-Inf*. <https://doi.org/10.3390/ijgi10080493>
4. Minka T et al. (2018) TrueSkill 2: an improved Bayesian skill rating system. Technical Report
5. Wang J (2021) Vision of China's future urban construction reform: In: the perspective of comprehensive prevention and control for multi disasters. *Sustainable cities and society*
6. Florida R, Rodríguez-Pose A, Storper M (2021) Cities in a post-COVID world. *Urban Studies*
7. Liu M, Jiang Y, He J (2021) Quantitative evaluation on street vitality: a case study of Zhoujiadu community in Shanghai. *Sustainability*

8. Lynch K (1964) *The image of the city*. MIT press
9. Kwok CYT, Wong MS, Chan KL, Kwan MP, Nichol JE, Liu CH, ... Kan Z (2021) Spatial analysis of the impact of urban geometry and socio-demographic characteristics on COVID-19, a study in Hong Kong. *Sci Total Environ*
10. Rode P, Floater G, Thomopoulos N, Docherty J, Schwinger P, Mahendra A, Fang W (2017) *Accessibility in cities: transport and urban form*. *Disrupting Mobility*
11. Bing X, Bingyu Z, Xiao X, Jingzhong L, Xiao X, Wanxia R (2020) A POI data-based study on urban functional areas of the resourcesbased city: a case study of Benxi, Liaoning. *Human Geogr*
12. Yan L (2015) *The progress of landscape planning: a brief discussion on urban parks*. *Chinese Science and technology periodical database engineering technology*

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

