



Examining the non-linear relationship between urban form and air temperature at street level: A case of Hong Kong

Lai Tian ^a, Tongping Hao ^a, Xinyu He ^b, Isabelle Chan ^c, Jianlei Niu ^d, P.W. Chan ^e, W.Y. Ng ^f, Jianxiang Huang ^{a,g,*}

^a Department of Urban Planning and Design, Faculty of Architecture, The University of Hong Kong, Pokfulam Rd, Hong Kong Special Administrative Region, China

^b Hong Kong Institute for Data Science, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon Tong, Hong Kong Special Administrative Region, China

^c Department of Real Estates and Construction, Faculty of Architecture, The University of Hong Kong, Pokfulam Rd, Hong Kong Special Administrative Region, China

^d Department of Building Environment and Energy Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong Special Administrative Region, China

^e The Hong Kong Observatory, 134A Nathan Road, Kowloon, Hong Kong Special Administrative Region, China

^f Environmental Protection Department, Hong Kong SAR Government, Hong Kong Special Administrative Region, China

^g Urban Systems Institute, The University of Hong Kong, Hong Kong Special Administrative Region, China



ARTICLE INFO

Keywords:

Urban warming

Street canyon

Urban form

Machine learning models

ABSTRACT

The relationship between urban form and localized air temperature has been studied extensively using linear regression. However, the findings remain inconsistent, and few studies have explored alternative data modeling techniques. With the rise of machine learning, there is an opportunity to explore new methods in urban climatology research. This study aims to test the hypothesis that machine learning models, rather than linear regression, can better explain and predict urban air temperature fluctuations in space and time. Measurements of air temperature were conducted at street level in Hong Kong. Urban form characteristics surrounding the measurement locations were extracted from geo-spatial databases and street view imagery. The datasets were then used to test the performance of linear regression, Artificial Neural Network (ANN), and Random Forest (RF) in predicting the spatial-temporal temperature fluctuations. The results indicate that the relationships between urban form and air temperature are predominantly non-linear. Both ANN and RF outperformed linear regression in prediction, with an MAE of 0.43 °C and 0.33 °C respectively. This study highlights the potentials of machine learning models in advancing knowledge of the impact of urban form on localized air temperature.

1. Introduction

In recent years, there has been a growing interest on surging urban air temperature amongst urban planners and researchers. The rise in urban temperature is influenced by urban form attributes such as the street canyon effect, alterations in wind patterns near tall buildings, and the release of heat from air-conditioning units and vehicles [1]. Conventional literature has typically employed a physics-based approach, using energy balance equations solved at various scales to predict temperature and airflow. Notable examples include Computational Fluid Dynamics (CFD), a highly effective tool. However, modeling a complex city using CFD entails a substantial computational burden, and researchers often encounter challenges from oversimplification to

validation [2]. To address this complexity, a data-driven approach has emerged recently, utilizing vast databases from sensor networks or simulations. Researchers extract urban form data from geo-spatial databases, such as LIDAR, street views, or remote sensing images, representing these areas with a set of continuous feature parameters. The relationships between temperature and urban form, usually expressed through regression coefficients, have been reported with implications for policy and planning practices [3,4].

While recent studies in urban heat research have utilized advanced non-linear techniques, these approaches typically rely on meteorological station data and remote sensing images to predict high-resolution land surface temperatures due to the requirement for extensive training data [5,6]. When it comes to predicting and explaining air

* Corresponding author. Department of Urban Planning and Design, Faculty of Architecture, The University of Hong Kong, Pokfulam Rd, Hong Kong Special Administrative Region, China.

E-mail address: jxhuang@hku.hk (J. Huang).

temperature fluctuations within street canyons from ground-truth data, most studies rely on linear regression or its variations, which assume a straight-line fit to the data using the least square method. However, few studies have examined their datasets to ensure they meet the underlying assumptions of least squares, which include linearity (i.e., a linear relationship between X and the mean of Y) and homoscedasticity (i.e., constant residual variance for any value of X). Moreover, there is often contradictory evidence regarding the direction and strength of the association. For example, some found that high-rise, high-density urban forms in Hong Kong contribute to extensive urban heat [7], while others reported the opposite within the same city [8,9]. Alternative data modeling techniques are rarely discussed, despite the advancement in non-linear models represented by machine learning algorithms, which have been applied to study building energy [10] and urban transport [11], and more recently in land surface temperature and urban heat island [12,13]. This limitation hinders the development of a more nuanced understanding and accurate prediction of street-level urban warming. There is a need to address the above gap and test non-linear models in this research domain.

The aim of this study is to test the hypothesis that non-linear models, such as machine learning algorithms, can provide a better explanation and prediction of urban air temperature fluctuations at street level in both space and time compared to traditional regression methods. The paper is structured around the following questions: 1) Do the relationships between air temperature and street canyon characteristics satisfy the linearity assumption required by regression analysis? 2) Can non-linear models, represented by machine learning algorithms, enhance our understanding and prediction of pedestrian air temperature in relation to urban form? To answer these questions, mobile and stationary temperature measurements were conducted on hundreds of streets in Hong Kong during the hot seasons. Urban form characteristics surrounding the measurement locations were extracted from geo-spatial databases and street view imagery. The datasets were then used to compare the performances of regression and non-linear models, including Artificial Neural Network and Random Forest. The distributions of mean air temperature across key parameters of urban form were examined, and the predictive power of linear and non-linear models were compared. The study concludes by discussing the implications of the findings for research and planning practices.

2. Literature review

The impact of urban form on localized air temperature in street canyon has been extensively studied. Most employ linear regression in data fitting, assuming a linear relationship between the two; while the direction and the strength of association vary widely from study to study, let alone their sample sizes and technical robustness. Advancement in non-linear models provides an alternative which may advance our understanding and predictive capacity of urban warming.

2.1. Urban form and street canyon temperature

Literature has associated localized air temperature with various attributes of street canyon characteristics. For instance, the Sky View Factor (SVF), a measure of the proportion of open sky visible at a particular location, is considered the most important determinant governing localized warming. However, findings from some of the most highly cited studies indexed by SCOPUS often contradict each other regarding the relationship between SVF and air temperature (Table 1). Many conclude that urban heat, measured variably using the air temperature, Mean Radiant Temperature (T_{mrt}), to Physiological Equivalent Temperature (PET), associate positively with SVF [7,14–20]; others yielded opposite findings, insisting negative associations between the two [8,21–24]. Some reported mixed or unclear relationships [25–29]. A majority of these studies employed linear regression, with some exceptions in the use of spatial regression, which is a derivative of the linear model. Although the discrepancies in finding may be influenced by other technical factors such as data aggregation and measurement period, the way these self-contradictory results are interpreted represents a challenge. Such discrepancies beg the question of whether the data modeling technique is part of this challenge, and whether the linearity assumption between urban warming and urban form, SVF included, hold in these conditions.

The height-to-width ratio (H/W), which measures the proportion of street canyons, is another popular metric for urban form. Evidence from Hong Kong suggest that H/W contributes positively to urban warming, while dense urban core with higher H/W values are between 1 and 1.5 °C warmer than those in the urban outskirts [9]. However, this finding was disputed by Deng et al. [25] in a study conducted in the central business district of Nanjing, a city sharing the humid subtropical

Table 1

A summary of relevant studies on urban heat and Sky View Factor, amongst others, using multi-variate regression techniques. The table reports only the regression coefficient of SVF.

Main Findings	Regression Coefficient of SVF on Heat (°C)	Data Modelling Technique	Dependent Variables	Sites	Study Location	Source:
Positive relationship	47.66	OLS Regression	Measured (T_{mrt} - T_a)	15	Curitiba, Brazil	[19]
	0.61–0.92	OLS Regression	PET	6	Taiwan	[17]
	20.41	OLS Regression	PET	6	Athens	[15]
	18.63–31.56	OLS Regression	T_{mrt}	17	Singapore	[20]
	34.16 for T_{mrt}	OLS Regression	T_{mrt} , PET	10	Hong Kong	[7]
	15.90 for PET					
	12.8 for T_a	OLS Regression	T_a , T_{mrt} , PET	17	Isfahan	[14]
	22.9 for T_{mrt}					
	13.31 for PET					
	0.39–1.02	Spatial regression	Land surface temperature	Multiple locations	Wuhan	[30]
Negative relationship	-0.71	OLS Regression	T_a	10	Lisbon	[21]
	-16	OLS Regression	T_{mrt}	6	Hong Kong	[8]
	~ -16.7 for T_{mrt}	OLS Regression	T_a , T_{mrt} , PET	Multiple in 4 urban areas	Montreal	[23]
	~ -17.5 for PET					
	-0.15	OLS Regression	T_{mrt}	Multiple along streets	Tokyo	[22]
Mixed or contradictory results	Insignificant	OLS Regression	T_a , PET	10	Tehran	[26]
	Insignificant	OLS Regression	T_{mrt}	22	Arizona	[29]
	Insignificant	OLS Regression	PET	24	Tehran	[27]
	Insignificant	OLS Regression	T_a , T_{mrt} , PET	12	Freiburg	[28]
	Insignificant for T_a , positive to T_{mrt}	OLS Regression				
	Contradictory results	t-test	Simulated T_a , T_{mrt} , PET	9	Nanjing	[25]
	Contradictory results	OLS Regression	T_a	26	Hong Kong	[9]

climate and the compact urban form with Hong Kong. They reported negative correlations between H/W and simulated air temperature, possibly explainable by building shadows cast in deep and narrow streets canyons. Similar observations were made by Darbani et al. [31] in a study in Mashhad, Iran. They concluded that H/W contributes negatively to localized air temperature, the mean radiant temperature, and the severity and duration of heat stress experienced by pedestrians.

Street orientation is another feature that influences localized air temperature. A study conducted by Abdollahzadeh and Biloria [32] in Sydney, Australia suggests that streets running north-south and north-west – southeast orientations have higher levels of thermal comfort compared with the rest, while east-west streets are the hottest, since it receives intense solar radiation throughout the day. Another study conducted in Wales, UK by Huang et al. [33] reached a different conclusion. They found that north-south oriented streets are more likely to experience heat stress during heatwaves, compared with those of the east-west oriented ones. The difference in findings above can be explained by the distinct climate, including solar angles and prevailing wind directions, between the UK and Australia.

The presence and quantity of greenery is also attributed to the cooling of air temperature on streets. Tree canopies shade pedestrians and building surfaces, reducing the solar heat gains from sunlight [34]. Trees also contribute to cooling through evapotranspiration, which absorbs latent heat from surrounding environment and potentially reduces building cooling energy consumption [35]. Although conflicting evidence were also reported in compact urban environments, where the cooling effect of greenery has been constrained by the limited planting surfaces and sunlight availability at the ground level [36].

2.2. Data modelling techniques

A majority of the data-driven studies in the field employ linear regression, a technique based on the least square method, i.e., of fitting multiple observations to a straight line by minimizing the sum of the squares of the residuals [19,26]. This approach, pioneered by statistician George [37] has become the preferred data fitting approach in economics, environmental and behavioral studies of the 20th century [38], and by extension, in the field of urban heat studies. Regressing temperature on urban form characteristics generate results that are easy to interpret. The relationships between dependent and independent variables, denoted as regression coefficients, can explain the magnitude and direction of the impact of X on Y, which is easy to comprehend by planners and policymakers. Regression models can also be used to predict localized air temperature, as exemplified by the Screening Tool for Estate Environment Evaluation (STEVE), a multi-variate linear regression model developed to predict localized urban air temperature using surrounding urban form characteristics [39].

Of the linear regression models in use, the most commonly used regression is the Ordinary Least Squares (OLS). Examples include the use of OLS regression to model the relationship between occupancy and equivalent temperature in three outdoor urban areas by Huang et al. [1]. Other regression techniques are also applied in the study of urban heat, suitable to the nature of data. Spatial regression, a variation of the linear regression with a linear additive specification, has been used to account for the spatial autocorrelation or the 'spillover' effect of urban heat in 2D space. For instance, You et al. [12] employed spatial regression to model the associations between UHI and particular urban form characteristics in Shanghai's new towns using Landsat thermography datasets. Another example by Chun and Guldmann [13] used the general spatial regression to model the UHI intensity. They concluded that open spaces, vegetation, building roof-top areas, and water strongly impact surface temperatures, and the general spatial model can achieve the best result amongst others, with implications for design and land-use policies. Other techniques, such as Logistic Regression, or Negative Binomial Regression, are used for count data, although both are regarded as generalized linear models. For instance, Hao et al. [40] used Negative

Binomial Regression to model the park attendance and geo-coded tweets in relation to the fluctuation of thermal environments, which has yielded a negative linear curve between the two.

However, recent studies have demonstrated the non-linear relationships between urban form characteristics and temperature. Due to the complexity of urban form, structures like buildings, roads, green-spaces, and water bodies exhibit differences in absorption, reflection, and heat dissipation. Traditional regression methods often fail to capture these intricate and spatially heterogeneous non-linear relationships. In contrast, machine learning algorithms can learn from large datasets to uncover subtle, non-intuitive interactions that may be missed by simpler linear methods. For example, a study by Gu et al. used XGBoost and SHAP algorithms shows that increased plot ratio has a positive impact on nighttime land surface temperatures in summer, while the contribution first rises then descends in the other three seasons (2024). Understanding and predicting such complex temperature fluctuations is crucial for accurately assessing and managing the urban thermal environment.

A predominant approach in ML studies in this field has been to utilize meteorological station data and remote sensing images for predicting high-resolution temperature profiles [41–43]. However, in dense urban areas, weather stations are often sparsely distributed and located in open areas within cities, making it challenging to capture the accurate air temperature within street canyons. Owing to the challenges in data acquisition, few studies have combined machine learning with in-situ measurement data to explore the effects of street canyon characteristics on pedestrian-level air temperatures.

2.3. Research gap

Despite the growing body of evidence in this field, significant research gaps remain two-fold. First, the validity of using linear regression to model the relationship between temperature and urban form characteristics has not been adequately examined. Most studies regard linear regression as norm, with assumed causalities either explicitly or tacitly between the dependent and independent variables [31], even if troubling signals were found which contradict the basic assumptions of linear regression. For example, Deng and Wong [25] discovered that the influence of Sky View Factor on air temperature diminishes when the Height-to-Width Ratio of street canyons exceeds a certain threshold. Similarly, Jusuf and Hien [39] found that the association between urban form and urban temperature can exhibit contrasting directions depending on the location and time. Despite these findings, the majority of researchers continue to rely on linear regression as their preferred data modeling technique.

Secondly, the literature in this field has paid little attention to alternative data modeling techniques beyond regression, despite significant advancements in machine learning algorithms. In a pioneering study, Chen et al. [44] explored the relationship between location-based air temperature and surrounding urban form characteristics using multiple machine learning algorithms. Their findings suggest that machine learning algorithms can achieve a reduction in the predicted root mean square error (RMSE) of approximately 0.2 °C compared to regression models. However, the implications for theory remain unclear. In the context of human body thermal comfort, machine learning has primarily been used to improve prediction accuracy at the expense of interpretability. For example, Hu et al. [45] found that machine learning outperformed the equations of the Predicted Mean Vote (PMV) by P. O. Fanger [46] by on average 13.1 %, but failed to explain how the underlying human heat physiology could be improved. In general, machine learning models are challenging to interpret and are often considered 'black box' in nature. They have yet to gain the same level of trust as linear regression models from the research community. The field of urban heat studies continues to be dominated by regression-based literature.

These above gaps underscore the need for a more critical

examination of the assumptions underlying linear regression and the exploration of alternative data modeling techniques. By doing so, one can enhance the understanding of the complex relationship between localized air temperature and urban form and improve the accuracy of predictions in support of sound design and policy.

3. Methods

In this study, machine learning models were developed to predict outdoor temperature from urban form parameters in Hong Kong. On-site temperature measurements of streets were conducted, and corresponding urban form parameters were obtained from Rhino models and Google Street View images. The importance of these parameters and temperature predictions were assessed using the Random Forest and Artificial Neural Network (ANN) models, while result interpretation was conducted using SHAP values analysis. The conceptual framework of the study is illustrated in Fig. 1.

3.1. Field measurement

Temperature profiles within street canyons in Hong Kong were measured using both mobile and stationary sensors. The reliability of the measured data surpasses that of simulated data, which often relies on simplified input parameters and assumptions. The measurements were carried out within segments of street canyons spanning a length of 13 km along the Hong Kong tramway, situated on the northern edge of Hong Kong Island. This area represents the historical urban center and is characterized by some of the highest built environment density worldwide. The street canyons in this region are confined by towering buildings, ranging from 40 to 60 stories in height, with a width of approximately 15–25 m from façade to façade [47]. These areas are known for extensive urban heat island effect [48], which poses a major concern for local urban planners [49].

Measurements were conducted at the street level using both mobile and stationary sensors. The mobile measurements were obtained from inside a double-level tram, which traverses the streets at an average speed of approximately 30 km/h. The sensors were located approximately 3 m above the ground along the road center line, with a distance of 10–20 m from building facades on both sides. The tram served as an ideal platform for measurements due to its natural ventilation, ensuring that the air temperature inside the tram closely matched the external conditions. Furthermore, since the trams operate on electricity, the potential influences from engine exhaust heat were minimized.

Additional measurements were obtained by walking, covering streets that were not accessible by tram and exhibiting varying widths and height profiles. Fig. 2 illustrates the locations of the mobile measurements taken from both the tram-mounted and pedestrian-mounted sensors. The measurement campaigns were conducted during the hot and moderate seasons, from June to November in 2023. Each measurement campaign lasted for a duration of 4–5 h, effectively capturing the seasonal and diurnal variations in temperature. Further details regarding the measurement schedule are included in Table 2.

The mobile sensors used in the study were equipped with components capable of capturing air temperature, humidity, and globe temperature. To prevent any influence from sunlight, the temperature and humidity sensors were placed inside a shading case and used to measure the relative humidity and air temperature, respectively. The black globe temperature sensor, on the other hand, was utilized to measure the radiant heat flux from the sky and urban surfaces, which was later used to compute the mean radiant temperature. The data collected by these sensors were uploaded and managed via HOBOlink, a web-based platform designed for sharing and managing environmental data (Table 3). The temperature sensors have an estimated response time of 5 min, which has been adjusted during subsequent data processing. Concurrent weather data were also collected from the nearest ground-based meteorological stations operated by the Hong Kong Observatory, namely Hong Kong Park, Happy Valley, and Shau Kei Wan (see Fig. 2). These data were used to represent the ambient meteorological conditions and were later utilized to compute the temperature differences, or the degree of temperature rise inside urban canyons.

In addition to the mobile sensors, stationary measurements were obtained from a roadside air quality monitoring station situated in Central (refer to Fig. 3). The sensors were situated on the station's rooftop, approximately 3 m above the ground. This station, operated by the Environmental Protection Department [50], was primarily designed to measure street-level air pollution. However, it also provided air temperature data, which played a critical role in calibrating the air pollution measurements. The air temperature data collected from the EPD were obtained continuously for a period spanning from Jun. 1 to Sept. 30 in 2023.

The Mean Radiant Temperature (T_{mrt}) was computed in order to crosscheck the robustness of air temperature measures. It represents the comprehensive average temperature of the radiation environment experienced by the human body. Considering the standard black globe (diameter = 0.15 m, emissivity = 0.95) utilized in the study, T_{mrt} can be calculated using formula (1) in accordance with Kuehn's formula [51].

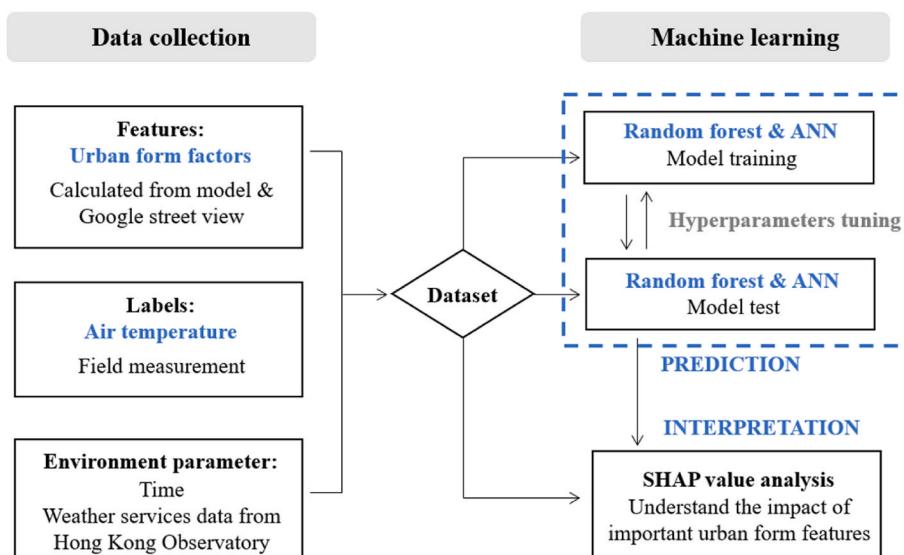


Fig. 1. Conceptual framework of the study.



Fig. 2. (a)Measurement route and weather stations on Hong Kong Island (source: HKO & Hong Kong Environmental Protection Department) (b)Equipment carried along to collect mobile data (c)Measure temperature from both tramways and pedestrians.

Table 2
Measurement time and parameters.

	Date	Measurement Period	Parameters
Mobile measurements	Jun. 1, 2023	11:35-16:10	Air Temperature
	Jun. 2, 2023	10:45-15:40	Relative Humidity
	Aug. 9, 2023	8:00-15:00	Black Globe Temperature
	Aug. 14, 2023	13:00-17:05	Dew Point
	Aug. 16, 2023	9:55-14:15	GPS Longitude & Latitude
	Nov. 4, 2023	10:00-13:00	Air Temperature
Stationary Measurements	Jun. 1, 2023 to Sept. 30, 2023	0:00-23:00	Air Temperature

Table 3
Sensors and equipment used in field studies.

Instrumentation	Measurement Parameter	Range	Accuracy	Manufacturer
	Black globe temperature	-40°C to 100°C	±0.2°C	HOBO Onset, Bourne, MA, USA
	Air temperature	-40°C to 75°C	±0.2°C	
	Relative humidity	0 % to 100 %	±2.5 %	
	Data logger (HOBO RX3000 station)			

$$T_{mrt} = \sqrt[4]{(T_g + 273.15)^4 + 2.5 \cdot 10^8 \cdot v_a^{0.6} \cdot (T_g - T_a)} - 273.15 \quad (1)$$

where T_g is the globe temperature ($^{\circ}\text{C}$), T_a is the air temperature($^{\circ}\text{C}$), and v_a is the localized wind speed, estimated by converting the wind speed recorded at Central Pier station to the pedestrian level in an urban environment.

3.2. Urban form characteristics

The urban form factors employed in this research include building coverage ratio, floor area ratio, average building height, street height-to-width (H/W) ratio, street width, street orientation and enclosure ratio, sky view and tree view factor, highest building within 50 m radius and distance from measurement point to the highest building. The selection of the variables was based on parameters found in existing studies [25, 52,53].

The 3D building geometries utilized in the study were obtained from the ib1000 digital map provided by the Hong Kong Lands Department [54]. These geometries were then processed in the Rhinoceros and ArcGIS software platforms. To extract the surrounding urban form characteristics at each measurement point, the GPS coordinates were geo-coded at a 1-min interval. The extraction process automatically collected the relevant information within a radius of 50 m from each measurement point. The selection of a 50 m radius threshold was determined to have the most significant impact on temperature compared to alternative values such as 20 m or 100 m [55]. The urban form characteristics encompassed various factors such as the average building height, building density, average height-to-width ratio, street orientation, greenery, and more (Fig. 4). To execute these steps, a computer script was developed and implemented using the Rhino-Grasshopper Plugin.

Sky View Factor (SVF) and Tree View Factor (TVF) are computed using image segmentations based on Google Street View data. Both are important parameters that describe the outdoor urban environment in terms of energy radiation transfer and the geometric relationship between different components of city streets. SVF and TVF respectively denote the geometric ratio between the number of visible sky and trees and the hemisphere enclosed by the horizontal plane from a given point



Fig. 3. Roadside monitoring station in Central (source: Google map street view).

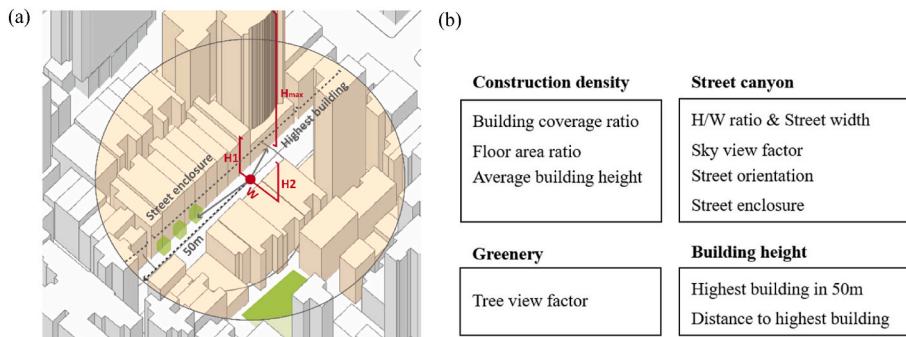


Fig. 4. (a) Features of one measuring point calculated within 50 m radius circle (b) Urban form factors.

[56].

The Google Street View (GSV)-based photographic method used for estimating SVF and TVF involves three main steps - downloading GSV panoramas from the Google database, applying image segmentation method to classify the sky and greenery region, and finally converting the cylindrical GSV panorama to a fisheye image for calculating SVF. Users can explore seamless street-level images of urban streets on Google Maps based on GSV panorama images. By inputting the geographic coordinates, all panoramas in the study area can be downloaded using the Google Street View Image API.

Deep learning frameworks were used to extract street features from these panorama images. Image segmentation based on deep convolutional neural networks (CNN) has been widely applied to image classification and pattern recognition. It allows a computational model made up of multiple processing layers to learn natural data and automatically

divide an image into different specific regions or objects of interest, including sky, trees, and buildings [57]. Pyramid Scene Parsing Network (PSPNet) is employed in this study. It uses a pre-trained semantic segmentation network based on the ADE20K dataset, which can predict 150 urban landscape item classes with an accuracy of 80.2 % [58]. Examples of image segmentation are presented in Fig. 5.

To reflect the shading effect of the surrounding buildings and vegetation at a given location, each segmented street view image was transformed into a fisheye image, with the upper half of the image above the horizon kept for the computation of the sky views. This step was implemented using the rectangular-to-polar coordinate transformation developed by Li et al [59], which can be expressed in Formula (2) to (6) below

$$C_x = C_y = W/2\pi \quad (2)$$

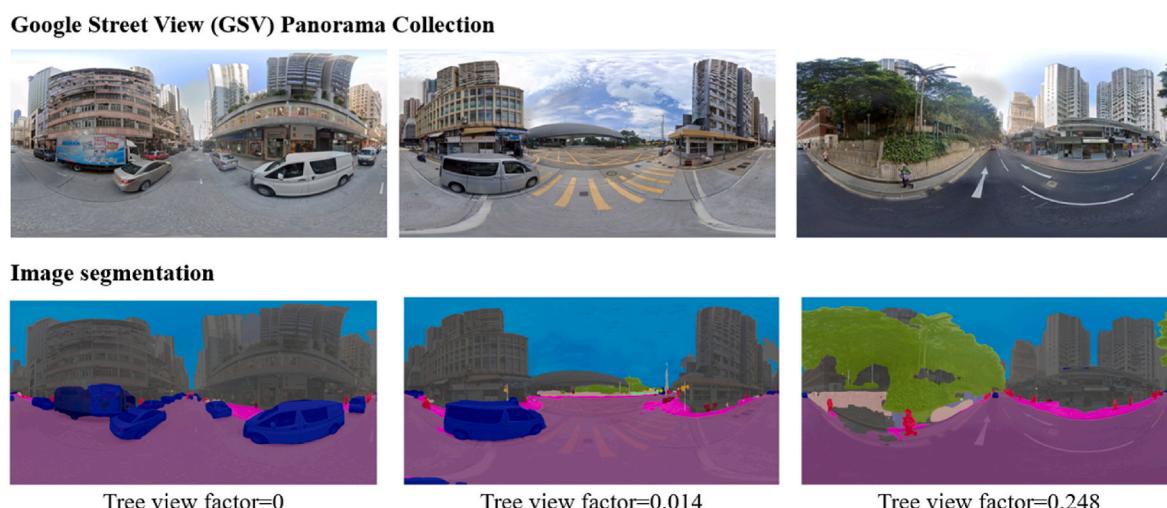


Fig. 5. Using image segmentation to identify key elements based on Google Street View.

$$x_p = \theta / 2\pi \bullet W \quad (3)$$

$$y_p = r / r_0 \bullet H \quad (4)$$

$$\theta = \begin{cases} \pi / 2 + \tan^{-1} \left(\frac{y_f - C_y}{x_f - C_x} x_f \right) & x_f < C_x \\ 3\pi / 2 + \tan^{-1} \left(\frac{y_f - C_y}{x_f - C_x} x_f \right) & x_f > C_x \end{cases} \quad (5)$$

$$r = \sqrt{(x_f - C_x)^2 + (y_f - C_y)^2} \quad (6)$$

where W and H are the width and height of the panorama image, respectively. Based on this, a fisheye image with a center at (C_x, C_y) and a radius of $W/2\pi$ was generated. The polar coordinates (r, θ) of each pixel in the fisheye image correspond to the rectangular coordinates (x_p, y_p) of each pixel in the panorama image. Fig. 6 presents the principle as well as examples of the transformation from panorama images to fisheye images.

3.3. Data modelling

Two commonly used machine learning algorithms, Random Forest and Artificial Neural Network, were used to model localized air temperature in comparison with linear regressions.

3.3.1. Machine learning algorithms

Both the Random Forest and Artificial Neural Network algorithms are used in this study. These algorithms are commonly used in machine learning literature and have been applied in relevant field to predict thermal comfort levels in indoor environments [60]. Random Forest is a non-linear machine learning technique that is constructed by averaging the predicted variables learned from many decision trees. Each tree selects a random sample of the full array as a bootstrap sample and trains independently during the splitting. The random selection of features for each tree during the splitting process adds an element of diversity, reducing bias and reducing the risk of overfitting [61]. Another significant benefit of random forest is its ability to assess the feature importance, which helps in identifying the most influential urban form factors for designers. Information gain is often used to measure the contribution of each feature to the accuracy of data partitioning [62].

Artificial Neural Network, a popular machine-learning algorithm, was used to model the non-linear relationships in this study. The algorithm consists of multiple layers of interconnected nodes, mimicking the neurons in a biological brain [63]. The input layer consists of a dataset

containing urban form parameters, time, and temperature data from Hong Kong Observatory, while the output layer is the measured temperature on-site. As the multiple interconnections are fully connected, the connection weights are changed after each epoch to minimize the mean absolute error between the real and predicted values. A z-score standardization process was applied to normalize the data across different magnitudes for both algorithms. To train the predictive models, Bayesian optimization was implemented to explore the optimal combination of machine learning hyperparameters in this study. Bayesian optimization utilizes Gaussian processes to represent the objective function and identifies parameter configurations that maximize the global performance [64]. The study defined five initial random search steps and 30 iterations of Bayesian optimization, with each iteration tested using five-fold cross-validation. In this study, the dataset is trained using 'Scikit-learn' package in Python.

Both Random Forest and Artificial Neural Network (ANN) were employed in this study to predict air temperature inside street canyons. The input datasets are from the measured air temperature data and surrounding urban form characteristics from the above steps. The date and time information were also included, which were transformed into solar zenith angles and azimuth angles based on the geographical coordinates the study location [65]. The Random Forest model's hyperparameters, including the number of trees, maximum depth, and maximum features, were optimized using Bayesian optimization and k-fold cross-validation. This iterative process was repeated to refine the search and identify the optimal hyperparameter configuration that delivers the best performance. The ANN model was trained in three separate steps in order to adjust for the uneven distribution of features for mobile and stationary data: 1) A complete ANN model was established to exclusively train the mobile data, capturing spatial features along with part of temporal features; 2) The weights of the hidden layers were frozen to avoid catastrophic forgetting, which occurs when the model forgets previously learned information while training on new tasks. Additional layers were then added above the frozen layers, and the stationary data were trained to learn the remaining temporal features; 3) All layers were unfrozen, and the entire model was fine-tuned with a very low learning rate, enabling the model to make small adjustments and refine its representations based on the combined knowledge acquired from both the mobile and stationary data. The models derived from the three-step approach were individually applied to the test set for evaluation. A 10 % random subset of the dataset was designated as the test dataset, while the remaining 90 % were split into training and validation sets following the 8:2 ratio in reference to published studies [66].

3.3.2. Interpretation of the non-linear relationship

The non-linear relationship between localized air temperature and

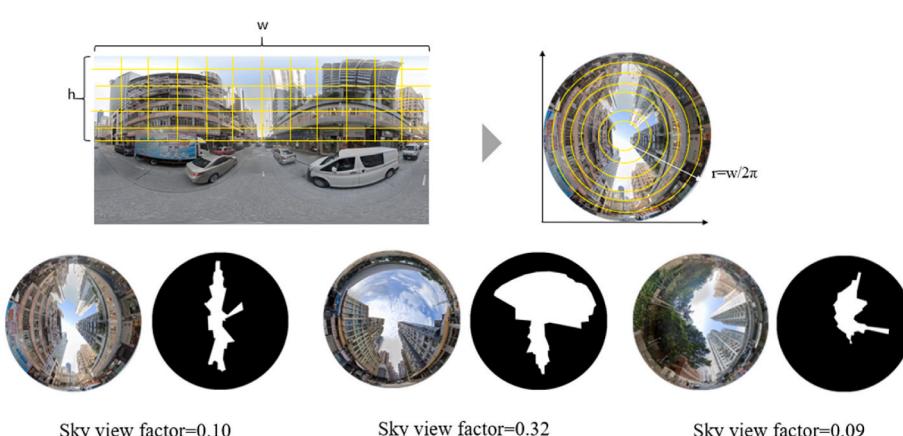


Fig. 6. Transforming panoramas to fisheye images to calculate SVF.

surrounding urban form has been interpreted using the SHapley Additive exPlanations (SHAP) value approach. Originally developed from the Cooperative Game Theory, SHAP has been used to decipher the ‘black-boxes’ of machine learning models [67] and recently in the study of human-environment interactive effect on thermal comfort [66]. SHAP values, a proxy of the “size-effect,” which represent the degree of association between air temperature and each urban form characteristics, were computed based on the marginal contribution of the latter amongst a large number of features in the input dataset. The contribution of each data feature’s SHAP value on the final output value (temperature rises) can be represented using Formula (7) below

$$g(z') = \phi_0 + \sum_{i=1}^M \phi_i z'_i \quad (7)$$

where $g(z')$ represents the explanatory model, z'_i is a binary vector denoting the presence (value of 1) or absence (value of 0) of a specific factor. The term ϕ_0 corresponds to the average value of the label, while ϕ_i represents the SHAP value. The calculation of SHAP values was implemented using the treeSHAP algorithm from the SHAP Python package, specifically designed for tree-based algorithms such as random forests and XGBoost. treeSHAP algorithm is considered reliable in the presence of data collinearity, i.e., two or more predictor variables are closely related to one another [66].

4. Results and discussion

4.1. Data characteristics

A dataset consisting of 4474 measured air temperature records was obtained, including 1545 from mobile sensors and 2929 from stationary sensors in the roadside monitoring station. Each data entry consists of the localized air temperature, the time stamp, the GPS coordinates, and surrounding urban form parameters.

The spatial-temporal variations of measured air temperature (red

line) in comparison with the ambient conditions (blue lines) were plotted in Fig. 7. A sizable difference can be observed between the two, with the former usually higher than the latter, especially during the afternoons and evenings. This disparity is 1.33°C during daytime hours during the study period, or up to 3°C during sunny and hot afternoons. A summary of data for both the mobile and stationary sensors are shown in Fig. 7 (a) and 7 (b) respectively. Both mobile and stationary observations indicate that in high-density urban districts, air temperatures may be higher than the local weather conditions. This could be attributed to the higher surface sensible heat flux and the intensified anthropogenic heat from traffic and air conditioner in street canyons. Additionally, the lower sky view factor in Hong Kong Island may also restrict radiative cooling during nighttime. Conversely, temperatures in street canyons from early morning to early afternoon tend to be comparable or occasionally lower than those recorded at the HKO stations. This finding can be explained by urban cool island result from building shading, higher thermal capacitance of the urban environment, or a deeper atmospheric boundary layer preventing the street canyons from warming up quickly in the early morning [68].

The Mean Radiant Temperature (T_{mrt}) exhibited sizable variations compared to the fluctuations of air temperature, particularly on warmer days (Fig. 8). This difference suggests that measured air temperature appeared unlikely to have been influenced by solar radiation, and the temperature sensors used in field measurement have been adequately protected inside the solar shield.

A summary of the data distribution of 11 urban form characteristics computed along the street canyons is provided in Fig. 9. The average street H/W ratio within the study area was 2.1, the average building height was about 56 m, and the average floor area ratio was 5.9. Moreover, the sky view factor was mostly distributed between 0 and 0.3, while the tree view factor was mostly below 0.3. These data reflect well the high-density and low-greening urban environment of Hong Kong. It is evident that parameters such as average building height, H/W ratio, street enclosure, and tree view factor do not adhere to a normal distribution, thus violating the underlying assumptions of linear regression in



Fig. 7. (a) Mobile measured data in street canyons in 6 days compared with data from three HKO stations; (b) Part of roadside monitoring station data in Central compared with data from three HKO stations.

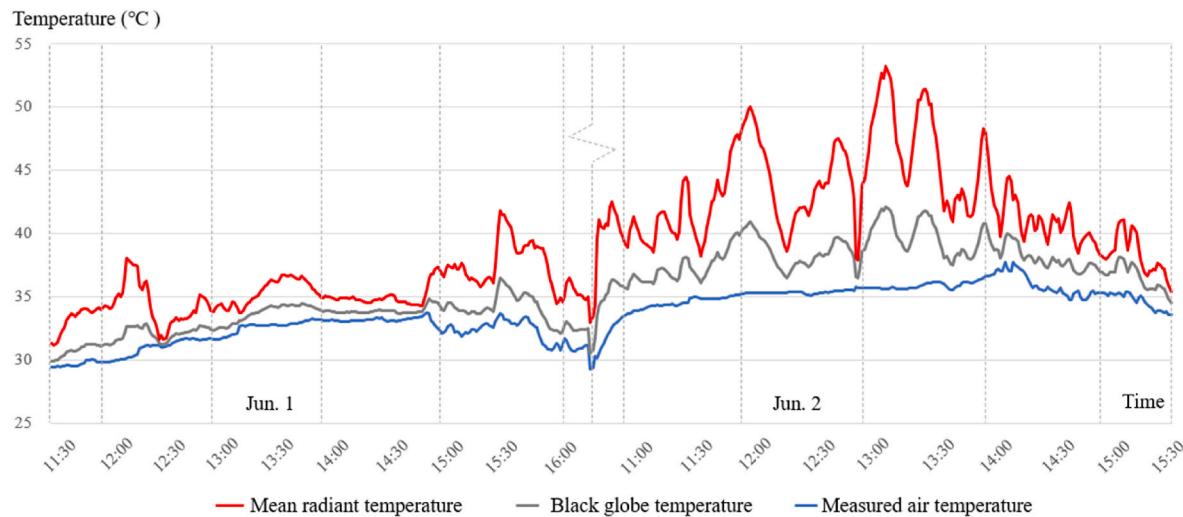


Fig. 8. Variation of measured Mean Radiant Temperature during two field study days.

general.

Air temperature measured using mobile sensors is mapped according to the GPS coordinates. Fig. 10 shows part of the results in Sheung Wan and Causeway Bay during the same time period in different days. Urban ‘hotspots,’ i.e., locations with consistent higher temperatures above the ambient, can be observed in places with less shading, or busy street intersections with extensive vehicular exhaust heat missions. The location of these ‘hotspots’ tend to stay consistent at different times of the day, suggesting the influences from surrounding urban form are also consistent.

4.2. The non-linear relationship between urban form and heat

The relationships between localized air temperature and urban form characteristics were found to be predominantly non-linear. The SHAP values, representing the contribution of building height, \bar{H}/W ratio, and SVF to ΔT , are plotted in Fig. 11.

The air cools in general as the average building height increases, suggesting that high-rise urban form can protect the overheating at street level, presumably by shading (Fig. 11 (a)). However, when the average building height exceeds 150 m, the increase in shaded areas becomes limited, resulting in a less pronounced cooling effect. Interestingly, the tendency of warming decreased first as the \bar{H}/W ratio increases, while it rises sharply as the urban warming increase sharply for street canyons with high \bar{H}/W ratio above 7 (Fig. 11 (b)). Our findings contradict with those reported in existing studies [25,69]. A possible explanation for the U-shaped curve is that the shading effect present in street canyons can cool the air at first, while in extremely narrow and deep urban canyons, the stagnant airflow and the radiative trapping of long-wave radiation can override the cooling benefits from shading. The relationship suggests the existence of an optimal \bar{H}/W ratio between 3 and 6, concerning the cooling of the street canyon as a goal. A similar pattern can also be observed in the influence of Sky View Factor (SVF) on the predicted temperature (Fig. 11 (c)), indicating that an SVF close to 0.27 is conducive to maximum cooling, while air temperature tends to rise in locations with either higher or lower values of SVF. This relationship cannot be captured by linear regression, which fits data a straight line.

Additionally, narrow streets (<40 m) and enclosed street facade (with enclosure factor in the range 0.6–0.9) were found to be conducive to cooler air, possibly due to the presence of shading from surrounding buildings and the reduction of radiative heat gains (Fig. 11(d) and (e)). Streets oriented to the North-South and Northeast-Southwest directions exhibit higher cooling potentials. This finding is in consistency with

existing literature [70]. Localized air tends to cool as urban greenery increases, measured in the green view factor. However, the magnitude of such cooling is minor, within 0.04 °C (Fig. 11 (g)). The finding is in consistency with published studies that the cooling effect provided by greenery in a high-density city is found to be lower than expected [40].

4.3. Relative contribution from urban form characteristics

The relative contributions of urban form characteristics in predicting urban heat, computed using the random forest algorithm, are summarized in Fig. 12, ranked by their importance with the total equaling one. The average building height within the 50 m radius is the most important factor barring the temporal influence, with some 13 % of contribution. The finding echos those from existing studies which suggest that tall buildings effectively reduce solar radiation received at the street level owing to the shadow they create, and this effect appears to be dominant [52,71]. The \bar{H}/W ratio is the second most important urban form factor with regard to temperature (~8 %), higher than those from SVF (~6 %). The finding contradicts with the majority of literature which regard SVF as the primary contributing factor [47]. This can be explained by the high-density environments in this study, in which SVF values are relatively low with smaller range of variations.

Other urban form factors are found to be less important than were reported in existing studies. The tree view factor accounts for less than 5 %, suggesting that urban vegetation may not be as important as it was found in landscape literature and policy documents [72,73], at least not in a high density urban environment such as Hong Kong. In fact, the finding of this paper echo a recent study also in Hong Kong, which suggests that urban vegetations are less effective for cooling purposes. The tree canopies provide little extra shade inside the long shadows casted by tall buildings; their evapotranspiration are also constrained due to the weak sunlight and the small ground-level surfaces available for planting [36]. Another low-importance factor is building coverage ratio, suggesting that the size of building footprint areas nearby matter less to urban heat in a high density city.

4.4. Predicting air temperature in street canyons

The performances of the ANN, Random Forest, and the OLS regression in predicting air temperature inside street canyons are compared with each other in Fig. 13. Both machine learning models, Random Forest and ANN, have performed reasonably well with low levels of prediction errors on the test dataset. The Random Forest model (Fig. 13 (a)) exhibited a Mean Absolute Error (MAE) of 0.33 °C and a R^2 value of

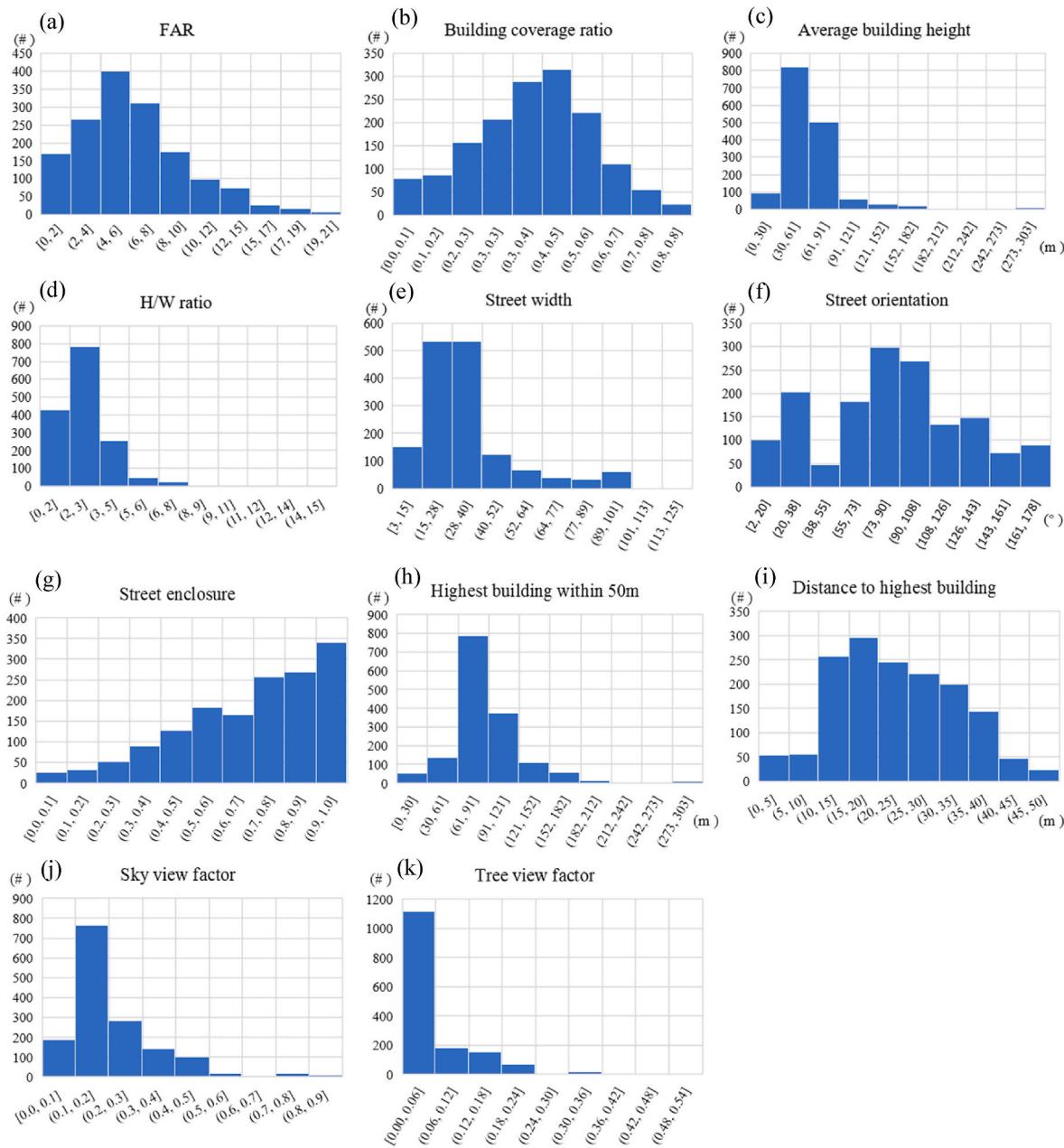


Fig. 9. Distribution of urban form characteristics.

0.96, while the ANN model scored 0.43 °C and 0.94 respectively (Fig. 13 (b)). In comparison, the OLS regression had an average error of 0.60 °C and RMSE of 0.78 °C on test dataset. Both the Random Forest and ANN models have outperformed the OLS regression by a significant margin.

A plausible explanation for the relatively robust prediction performances of machine learning models is their non-linear nature, which may be less vulnerable to uncertainties brought by overfitting and multicollinearity compared with regression. Prediction errors measured in MAE, RMSE, and R² on the training, validation and test dataset of both the Random Forest and ANN are listed in Table 4. RF has the least bias and lower model complexity, while ANN has the least variations across the training, validation and test set, suggesting the unlikelihood of overfitting. An ideal model should have low bias, low difference between training and test score, and reduced model complexity. We do not claim that machine learning is the perfect model to meet all these requirements, but based on the results, it performs better than linear regression in terms of RMSE and R². This is probably due to that machine

learning model typically capture the multidimensional influence of urban form, rather than focusing on the weight of each individual feature. They may be less susceptible to multicollinearity compared with linear regression.

Furthermore, the machine learning models can achieve stable prediction performances with a reasonable sample size of several thousands of data entries. For different sample sizes of Random Forest model, a 10 % random subset of each dataset was designated as the test dataset, and the model performance is plotted in Fig. 14. As the sample size increases, the model's R² increases in general, while the MAE and RMSE decrease steadily, suggesting improvement of fitness degree and decrease of error. These values however have largely stabilized when the sample size exceeds 3,000, indicating that stable prediction performances can be expected with a finite sample.



Fig. 10. The temperature difference (ΔT) between measured data and weather services data (a) on Aug. 14, 2023 in Causeway Bay, (b) on Aug. 16, 2023 in Causeway Bay, (c) on Jun. 1, 2023 in Sheung Wan, and (d) on Jun. 2, 2023 in Sheung Wan.

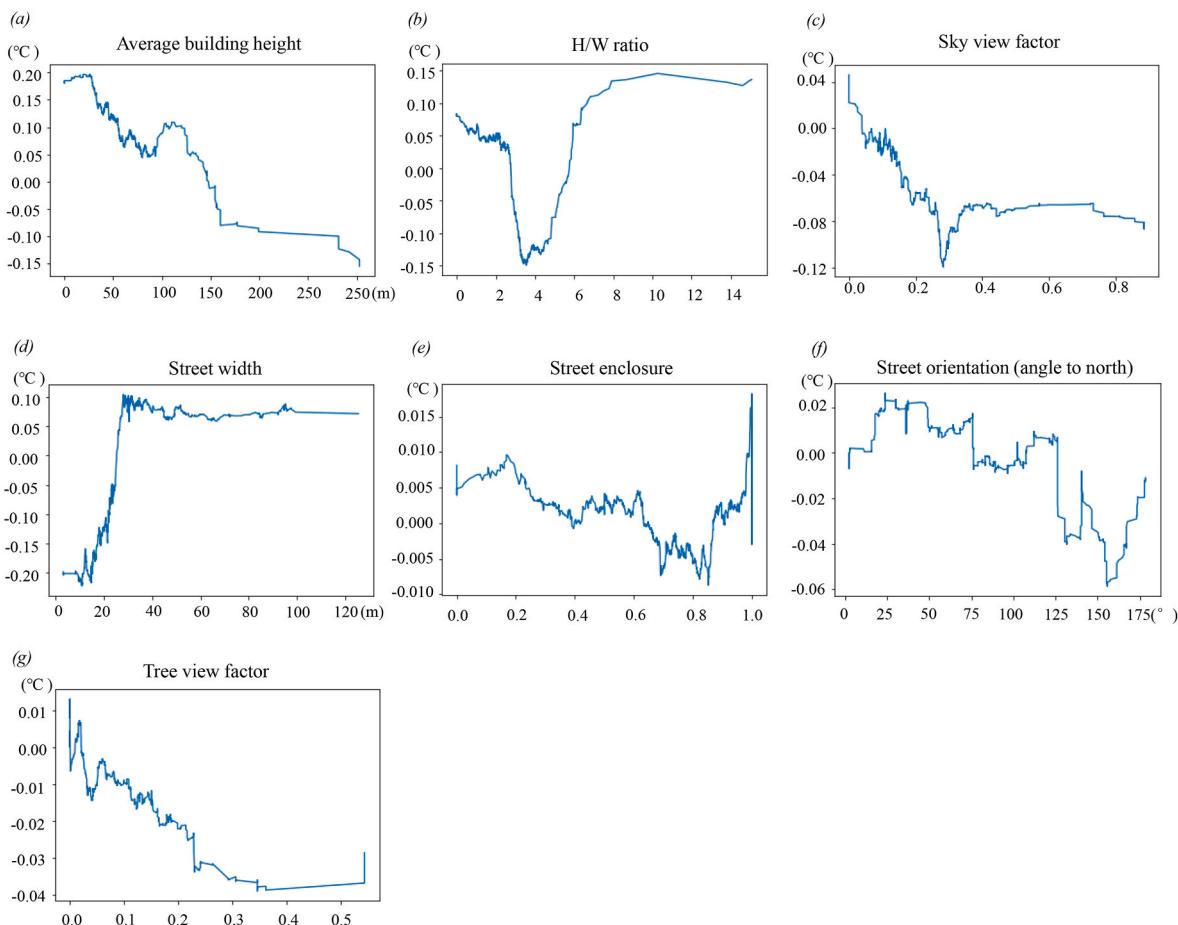


Fig. 11. The influence of individual urban form characteristic on predicted air temperature difference using SHAP analysis.

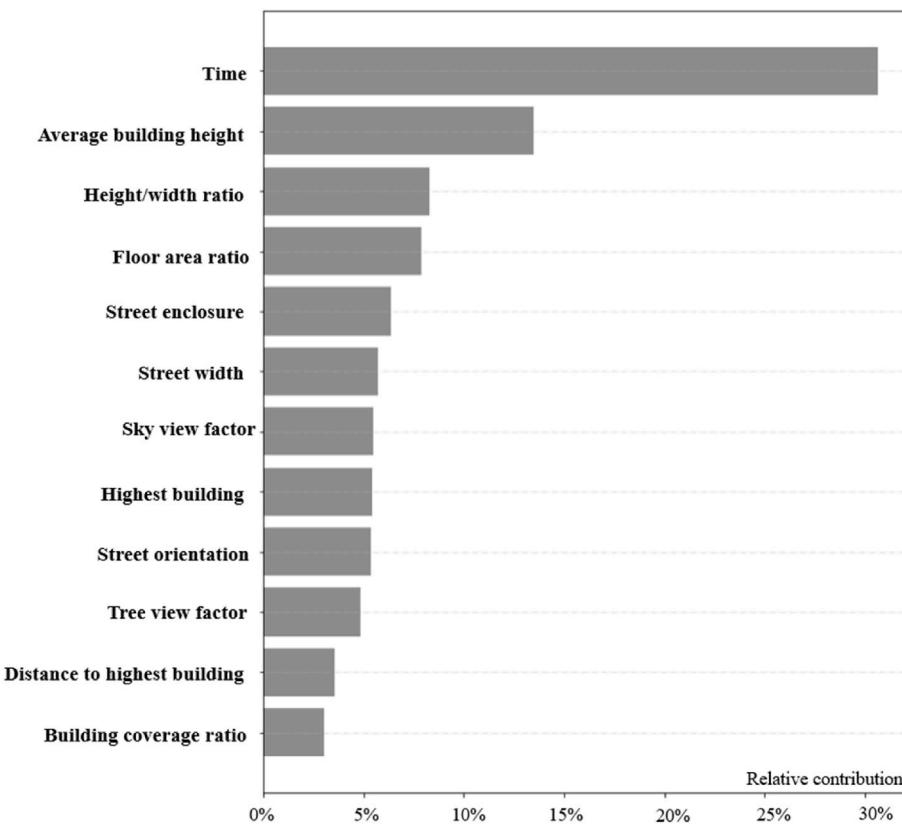


Fig. 12. The Importance of urban form characteristics in contribution to warming of the air inside street canyons (ΔT).

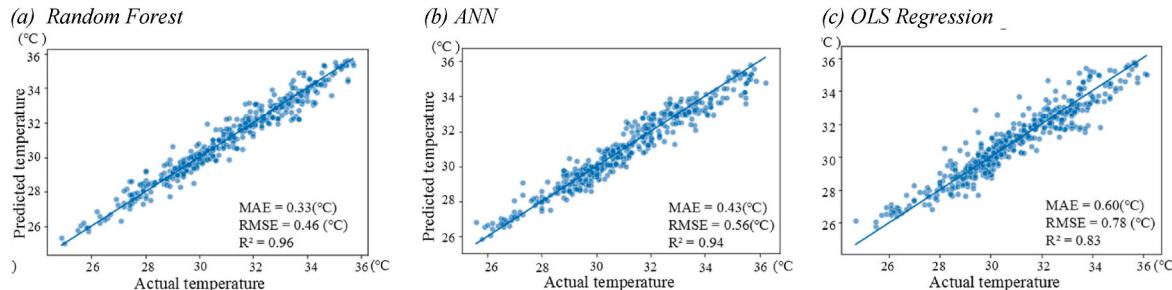


Fig. 13. Prediction performance of Random Forest (a) ANN (b), and OLS regression (c) on the test dataset.

Table 4
Comparison of accuracy and fitting degree of different non-linear models.

		Random Forest	ANN	OLS Regression
Training dataset	MAE	0.12	0.40	0.54
	RMSE	0.17	0.53	0.73
	R ²	0.99	0.94	0.90
Validation dataset	MAE	0.21	0.41	n.a.
	RMSE	0.30	0.55	n.a.
	R ²	0.97	0.94	n.a.
Test dataset	MAE	0.33	0.43	0.60
	RMSE	0.46	0.56	0.78
	R ²	0.96	0.94	0.83

4.5. Discussion

A key takeaway for the research community is the imperative to reflect upon the use of linear regression as a norm in the study of urban heat and urban form. Evidence obtained above suggest that the relationship between the two is predominantly non-linear (Fig. 11).

Furthermore, the distribution of residuals along the range of variables violated the homoscedasticity assumption, that is, the variation of error terms depends on the value of the predictor variables. Fig. 15 shows the standardized residuals of the OLS regression (a) and the Random Forest model (b) plotted along the air temperature using the test dataset obtained from mobile sensors. For the OLS regression, the residuals cannot be regarded as evenly distributed around zero on continuous independent variables. In contrast, the residuals for the Random Forest model are more evenly distributed (Fig. 15 (b)). The above evidence suggests that the premises in support of linear regression to analyze localized air temperature cannot be guaranteed, thus, the conditions can no longer satisfy the Gauss-Markov Theorem that the OLS is the best estimator with the lowest sampling variance [74]. Taking linear regression for granted without questioning its suitability can lead to irrelevant theory and questionable conclusions, according to statistician Leo Breiman [61]. It may also mislead policy and design, as demonstrated by the conflicting findings in the literature review (Table 1). There is a responsibility for the research community to robustly check the assumptions of the data modelling techniques of choice.

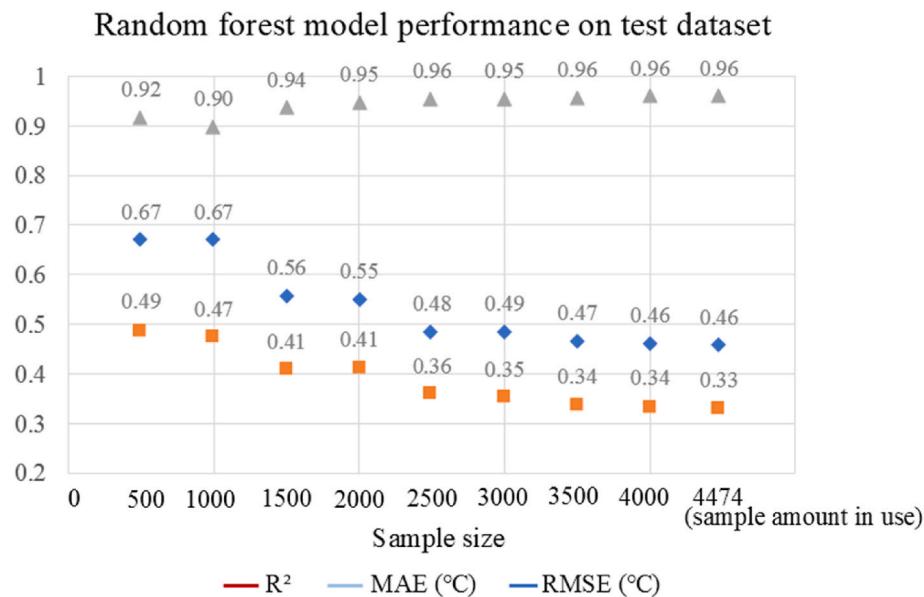


Fig. 14. Prediction accuracy tends to be stable as sample size increases.

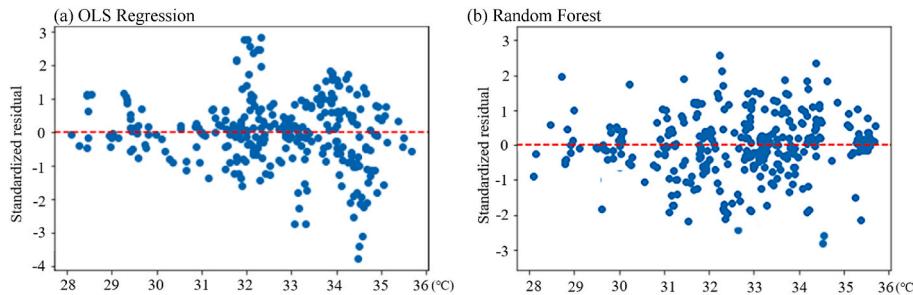


Fig. 15. Residual plot of (a) OLS regression and (b) random forest model for the test dataset from mobile sensors.

The practical message is that machine learning algorithms offer an alternative to predict and interpret the complex relationships between air temperature and urban form. Machine learning models exhibit significantly higher predictive power compared to OLS regression models, as illustrated in (Fig. 13). Both Random Forest and ANN provide a more comprehensive understanding of localized urban warming, surpassing the OLS regression. Importantly, machine learning models are not "black boxes" anymore, as they can effectively delineate the contribution of each urban form characteristic to temperature surge, as demonstrated in this study (Fig. 11). These contributions can be as intuitive as regression coefficients, and it is reasonable to expect that machine learning models will gradually gain the same level of trust as linear regression models currently enjoy within the research community. This advantage positions machine learning as a reliable tool for assessing heat risk in cities, supporting the development of heat-resilient policies and urban planning initiatives.

The study is limited in several aspects. The measurement data has been obtained from Hong Kong Island. The street canyons measured are mostly running in the east-west direction, with an average width between 20 and 40 m. The urban form attributes obtained may not represent other urban districts in Kowloon or the New Territories, let alone other cities. The machine learning models and findings developed in this study, therefore, are not intended to be used outside of the Hong Kong Island. This limited transferability can be a common challenge for machine learning models trained on a localized dataset. The next step is to collect more data from urban environments with more diverse conditions, with various density, street configurations, and traffic

conditions. Additionally, future research should include an extensive analysis of various land cover types, including asphalt pavement, impervious paving, grassland, trees, and water bodies, etc. These enhancements will allow for a more comprehensive understanding of the complex interactions between urban characteristics and their impact on air temperature patterns.

5. Conclusion

This study has investigated the impact of urban form on localized air temperature, with the hypothesis that non-linear models such as Random Forest and ANN enjoy an advantage in both explanation and prediction, compared with OLS regression. A relatively large dataset consisting of measured temperature were obtained and analyzed in Hong Kong. The findings confirmed the hypothesis. Random Forest and ANN achieved a mean absolute error of 0.33 °C and 0.43 °C respectively, both lower than those of the OLS regression. The relationship between air temperature and surrounding urban form conditions was found to be predominantly non-linear. The relationship between H/W ratio, SVF, and temperature exhibited a U-shaped curve, which cannot be captured by linear regressions. The key takeaway for the research community is to reflect upon the use of linear regression as a norm in the study of urban heat and urban form, and the practical message is that machine learning algorithms offer a powerful alternative. Findings of this study can inform design and policy to mitigate the temperature surge at street level in a high-density city.

CRediT authorship contribution statement

Lai Tian: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis. **Tongping Hao:** Validation, Project administration, Investigation. **Xinyu He:** Validation, Software, Methodology. **Isabelle Chan:** Writing – review & editing. **Jianlei Niu:** Writing – review & editing, Supervision, Resources. **P.W. Chan:** Data curation. **W.Y. Ng:** Data curation. **Jianxiang Huang:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

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.

Data availability

Data will be made available on request.

Acknowledgements

The research is supported by two grants from the National Natural Science Foundation of China (#51978594 & #51708473). The study is also funded in part by the Hong Kong Research Grants Council Theme-Based Research Scheme under Grant T22-504/21-R. The authors thank Mr. Xu Tang, Mr. Jinglei Li, and Mr. Yiqi Liu for their able assistance.

References

- [1] J. Huang, P. Jones, A. Zhang, R. Peng, X. Li, P. Chan, Urban building energy and climate (UrBEC) simulation: example application and field evaluation in sai ying pun, Hong Kong, Energy Build. 207 (2020) 109580, <https://doi.org/10.1016/j.enbuild.2019.109580>.
- [2] P.A. Mirzaei, CFD modeling of micro and urban climates: problems to be solved in the new decade, Sustain. Cities Soc. 69 (2021) 102839, <https://doi.org/10.1016/j.scs.2021.102839>.
- [3] A.M. Hassan, N.A. Megahed, COVID-19 and urban spaces: a new integrated CFD approach for public health opportunities, Build. Environ. 204 (2021) 108131, <https://doi.org/10.1016/j.buildenv.2021.108131>.
- [4] J. Huang, C. Yang, H. Tongping, P. Jones, Heat stress and outdoor activities in open spaces of public housing estates in Hong Kong: a perspective of the elderly community, Indoor Built Environ. 31(6) (2020) 1447–1463, <https://doi.org/10.1177/1420326X20950448>.
- [5] X. Gu, Z. Wu, X. Liu, R. Qiao, Q. Jiang, Exploring the nonlinear interplay between urban morphology and nighttime thermal environment, Sustain. Cities Soc. 101 (2024) 105176, <https://doi.org/10.1016/j.scs.2024.105176>.
- [6] Y. Guo, J. Unger, A. Khabibolla, G. Tian, R. He, H. Li, T. Gál, Modeling urban air temperature using satellite-derived surface temperature, meteorological data, and local climate zone pattern—a case study in Szeged, Hungary, Theor. Appl. Climatol. 155 (5) (2024) 3841–3859, <https://doi.org/10.1007/s00704-024-04852-7>.
- [7] T.E. Morakinyo, W. Ouyang, K.K.-L. Lau, C. Ren, E. Ng, Right tree, right place (urban canyon): tree species selection approach for optimum urban heat mitigation - development and evaluation, Sci. Total Environ. 719 (2020) 137461, <https://doi.org/10.1016/j.scitotenv.2020.137461>.
- [8] A. Lai, M. Maing, E. Ng, Observational studies of mean radiant temperature across different outdoor spaces under shaded conditions in densely built environment, Build. Environ. 114 (2017) 397–409, <https://doi.org/10.1016/j.buildenv.2016.12.034>.
- [9] R. Giridharan, S. Ganeshan, S.S.Y. Lau, Daytime urban heat island effect in high-rise and high-density residential developments in Hong Kong, Energy Build. 36 (6) (2004) 525–534, <https://doi.org/10.1016/j.enbuild.2003.12.016>.
- [10] C. Robinson, B. Dilkina, J. Hubbs, W. Zhang, S. Guhathakurta, M.A. Brown, R. M. Pendyala, Machine learning approaches for estimating commercial building energy consumption, Appl. Energy 208 (2017) 889–904, <https://doi.org/10.1016/j.apenergy.2017.09.060>.
- [11] C. Ding, X. Cao, Jason, P. Næss, Applying gradient boosting decision trees to examine non-linear effects of the built environment on driving distance in Oslo, Transport. Res. Pol. Pract. 110 (2018) 107–117, <https://doi.org/10.1016/j.trap.2018.02.009>.
- [12] M. You, J. Huang, C. Guan, Are new towns prone to urban heat island effect? Implications for planning form and function, Sustain. Cities Soc. 99 (2023) 104939, <https://doi.org/10.1016/j.scs.2023.104939>.
- [13] B. Chun, J.M. Guldmann, Spatial statistical analysis and simulation of the urban heat island in high-density central cities, Landsc. Urban Plann. 125 (2014) 76–88, <https://doi.org/10.1016/j.landurbplan.2014.01.016>.
- [14] A. Ahmad Venhari, M. Templierk, M. Taleghani, The role of sky view factor and urban street greenery in human thermal comfort and heat stress in a desert climate, J. Arid Environ. 166 (2019) 68–76, <https://doi.org/10.1016/j.jaridenv.2019.04.009>.
- [15] I. Charalampopoulos, I. Tsilos, A. Chronopoulou-Sereli, A. Matzarakis, Analysis of thermal bioclimate in various urban configurations in Athens, Greece, Urban Ecosyst. 16 (2) (2013) 217–233, <https://doi.org/10.1007/s11252-012-0252-5>.
- [16] X. He, S. Miao, S. Shen, J. Li, B. Zhang, Z. Zhang, X. Chen, Influence of sky view factor on outdoor thermal environment and physiological equivalent temperature, Int. J. Biometeorol. 59 (3) (2015) 285–297, <https://doi.org/10.1007/s00484-014-0841-5>.
- [17] R.-L. Hwang, T.-P. Lin, A. Matzarakis, Seasonal effects of urban street shading on long-term outdoor thermal comfort, Build. Environ. 46 (4) (2011) 863–870, <https://doi.org/10.1016/j.buildenv.2010.10.017>.
- [18] E. Krüger, P. Drach, P. Bröde, Outdoor comfort study in Rio de Janeiro: site-related context effects on reported thermal sensation, Int. J. Biometeorol. 61 (2016), <https://doi.org/10.1007/s00484-016-1226-8>.
- [19] E.L. Krüger, F.O. Minella, F. Rasia, Impact of urban geometry on outdoor thermal comfort and air quality from field measurements in Curitiba, Brazil, Build. Environ. 46 (3) (2011) 621–634, <https://doi.org/10.1016/j.buildenv.2010.09.006>.
- [20] C.L. Tan, N.H. Wong, S.K. Jusuf, Outdoor mean radiant temperature estimation in the tropical urban environment, Build. Environ. 64 (2013) 118–129, <https://doi.org/10.1016/j.buildenv.2013.03.012>.
- [21] H. Andrade, M.-J. Alcoforado, Microclimatic variation of thermal comfort in a district of Lisbon (Telheiras) at night, Theor. Appl. Climatol. 92 (3) (2008) 225–237, <https://doi.org/10.1007/s00704-007-0321-5>.
- [22] J.K. Vanos, E. Kosaka, A. Iida, M. Yokohari, A. Middel, I. Scott-Fleming, R. D. Brown, Planning for spectator thermal comfort and health in the face of extreme heat: the Tokyo 2020 Olympic marathons, Sci. Total Environ. 657 (2019) 904–917, <https://doi.org/10.1016/j.scitotenv.2018.11.447>.
- [23] Y. Wang, H. Akbari, Effect of sky view factor on outdoor temperature and comfort in montreal, Environ. Eng. Sci. 31 (6) (2014) 272–287, <https://doi.org/10.1089/ees.2013.0430>.
- [24] T. Xi, Q. Li, A. Mochida, Q. Meng, Study on the outdoor thermal environment and thermal comfort around campus clusters in subtropical urban areas, Build. Environ. 52 (2012) 162–170, <https://doi.org/10.1016/j.buildenv.2011.11.006>.
- [25] J.-Y. Deng, N.H. Wong, Impact of urban canyon geometries on outdoor thermal comfort in central business districts, Sustain. Cities Soc. 53 (2020) 101966, <https://doi.org/10.1016/j.scs.2019.101966>.
- [26] H. Farhadi, M. Faizi, H. Sanaieian, Mitigating the urban heat island in a residential area in Tehran: investigating the role of vegetation, materials, and orientation of buildings, Sustain. Cities Soc. 46 (2019) 101448, <https://doi.org/10.1016/j.scs.2019.101448>.
- [27] A. Karimi, H. Sanaieian, H. Farhadi, S. Norouzian-Maleki, Evaluation of the thermal indices and thermal comfort improvement by different vegetation species and materials in a medium-sized urban park, Energy Rep. 6 (2020) 1670–1684, <https://doi.org/10.1016/j.egyr.2020.06.015>.
- [28] H. Lee, H. Mayer, D. Schindler, Importance of 3-D radiant flux densities for outdoor human thermal comfort on clear-sky summer days in Freiburg, Southwest Germany, Meteorol. Z. 23 (3) (2014) 315–330, <https://doi.org/10.1127/0941-2948/2014/0536>.
- [29] A. Middel, E.S. Krayenhoff, Micrometeorological determinants of pedestrian thermal exposure during record-breaking heat in Tempe, Arizona: introducing the MaRTy observational platform, Sci. Total Environ. 687 (2019) 137–151, <https://doi.org/10.1016/j.scitotenv.2019.06.085>.
- [30] C. Yin, M. Yuan, Y. Lu, Y. Huang, Y. Liu, Effects of urban form on the urban heat island effect based on spatial regression model, Sci. Total Environ. 634 (2018) 696–704, <https://doi.org/10.1016/j.scitotenv.2018.03.350>.
- [31] E.S. Darbani, D.M. Parapari, J. Boland, E. Sharifi, Impacts of urban form and urban heat island on the outdoor thermal comfort: a pilot study on Mashhad, Int. J. Biometeorol. 65 (10) (2021) 1–17, <https://doi.org/10.1007/s00484-021-02091-3> Published.
- [32] N. Abdollahzadeh, N. Biloria, Outdoor thermal comfort: analyzing the impact of urban configurations on the thermal performance of street canyons in the humid subtropical climate of Sydney, Frontiers Arch. Res. 10 (2) (2021) 394–409, <https://doi.org/10.1016/j.foar.2020.11.006>.
- [33] J. Huang, X. Tang, P. Jones, T. Hao, R. Tundokova, C. Walmsley, S. Lannon, P. Frost, J. Jackson, Mapping pedestrian heat stress in current and future heatwaves in cardiff, newport, and wrexham in Wales, UK, Build. Environ. 251 (2024), <https://doi.org/10.1016/j.buildenv.2024.111168>.
- [34] H. Akbari, D.M. Kurn, S.E. Bretz, J.W. Hanford, Peak power and cooling energy savings of shade trees, Energy Build. 25 (2) (1997) 139–148, [https://doi.org/10.1016/S0378-7788\(96\)01003-1](https://doi.org/10.1016/S0378-7788(96)01003-1).
- [35] N.H. Wong, C.L. Tan, D.D. Kolokotsa, H. Takebayashi, Greenery as a mitigation and adaptation strategy to urban heat, Nat. Rev. Earth Environ. 2 (3) (2021) 166–181, <https://doi.org/10.1038/s43017-020-00129-5>.
- [36] J. Huang, T. Hao, Y. Wang, P. Jones, A street-scale simulation model for the cooling performance of urban greenery: evidence from a high-density city, Sustain. Cities Soc. 82 (2022) 103908, <https://doi.org/10.1016/j.scs.2022.103908>.
- [37] U. Yule, On the theory of correlation for any number of variables, treated by a new system of notation, Proc. R. Soc. Lond. - Ser. A Contain. Pap. a Math. Phys. Character 79 (529) (1907) 182–193, <https://doi.org/10.1098/rspa.1907.0028>.

- [38] S.M. Stigler, The history of statistics: the measurement of uncertainty before 1900, in: *Contemporary Sociology*, 16, Belknap Press, 1987, <https://doi.org/10.2307/2069855>.
- [39] S.K. Jusuf, W.N. Hien, *Development of Empirical Models for an Estate Level Air Temperature Prediction in Singapore*, 2009.
- [40] T. Hao, H. Chang, L. Sisi, P. Jones, P.W. Chan, L. Li, J. Huang, Heat and park attendance: evidence from 'small data' and 'big data' in Hong Kong, *Build. Environ.* 234 (2023), <https://doi.org/10.1016/j.buildenv.2023.110123>.
- [41] X. Ding, Y. Zhao, Y. Fan, Y. Li, J. Ge, Machine learning-assisted mapping of city-scale air temperature: using sparse meteorological data for urban climate modeling and adaptation, *Build. Environ.* 234 (2023) 110211, <https://doi.org/10.1016/j.buildenv.2023.110211>.
- [42] A. Hassani, G.S. Santos, P. Schneider, N. Castell, Interpolation, satellite-based machine learning, or meteorological simulation? A comparison analysis for spatio-temporal mapping of mesoscale urban air temperature, *Environ. Model. Assess.* 29 (2) (2024) 291–306, <https://doi.org/10.1007/s10666-023-09943-9>.
- [43] J.W. Oh, J. Ngarambe, P.N. Duhirwe, G.Y. Yun, M. Santamouris, Using deep-learning to forecast the magnitude and characteristics of urban heat island in Seoul Korea, *Sci. Rep.* 10 (1) (2020) 3559, <https://doi.org/10.1038/s41598-020-60632-z>.
- [44] S. Chen, N.H. Wong, W. Zhang, M. Ignatius, The impact of urban morphology on the spatiotemporal dimension of estate-level air temperature: a case study in the tropics, *Build. Environ.* 228 (2023) 109843, <https://doi.org/10.1016/j.buildenv.2022.109843>.
- [45] W. Hu, Y. Wen, K. Guan, G. Jin, K. Tseng, iTCM: toward learning-based thermal comfort modeling via pervasive sensing for smart buildings, *IEEE Internet Things J.* (2018) 1, <https://doi.org/10.1109/JIOT.2018.2861831>.
- [46] P.O. Fanger, *Thermal Comfort*, Danish Technical Press, 1970.
- [47] L. Chen, E. Ng, X. An, C. Ren, M. Lee, U. Wang, Z. He, Sky view factor analysis of street canyons and its implications for daytime intra-urban air temperature differentials in high-rise, high-density urban areas of Hong Kong: a GIS-based simulation approach, *Int. J. Climatol.* 32 (1) (2012) 121–136, <https://doi.org/10.1002/joc.2243>.
- [48] F. Ali Touider, H. Mayer, Effects of asymmetry, galleries, overhanging façades and vegetation on thermal comfort in urban street canyons, *Solar Energy - SOLAR ENERG* 81 (2007) 742–754, <https://doi.org/10.1016/j.solener.2006.10.007>.
- [49] E. Ng, V. Cheng, Urban human thermal comfort in hot and humid Hong Kong, *Energy Build.* 55 (2012) 51–65, <https://doi.org/10.1016/j.enbuild.2011.09.025>.
- [50] Hong Kong EPD, Air quality monitoring stations info. <https://www.aqhi.gov.hk/en/monitoring-network/air-quality-monitoring-stationsf9dd.html?stationid=79>, 2023. (Accessed 31 January 2024).
- [51] L.A. Kuehn, R.A. Stubbs, R.S. Weaver, Theory of the globe thermometer, *J. Appl. Physiol.* 29 (5) (1970) 750–757, <https://doi.org/10.1152/jappl.1970.29.5.750>.
- [52] E. Jamei, P. Rajagopalan, M. Seyedmahmoudian, Y. Jamei, Review on the impact of urban geometry and pedestrian level greening on outdoor thermal comfort, *Renew. Sustain. Energy Rev.* 54 (2016) 1002–1017, <https://doi.org/10.1016/j.rser.2015.10.104>.
- [53] S. Mokhtar, C. Reinhart, Towards scalable and actionable pedestrian outdoor thermal comfort estimation: a progressive modelling approach, *Build. Environ.* 242 (2023) 110547, <https://doi.org/10.1016/j.buildenv.2023.110547>.
- [54] Lands Department, Hong Kong digital map. <https://www.landsd.gov.hk/en/survey-mapping/mapping/multi-scale-topographic-mapping/digital-map.html>, 2021. (Accessed 22 January 2024).
- [55] H. Jin, P. Cui, N.H. Wong, M. Ignatius, Assessing the effects of urban morphology parameters on microclimate in Singapore to control the urban heat island effect, *Sustainability* 10 (1) (2018), <https://doi.org/10.3390/su10010206>.
- [56] T.R. Oke, Canyon geometry and the nocturnal urban heat island: comparison of scale model and field observations, *J. Climatol.* 1 (3) (1981) 237–254, <https://doi.org/10.1002/joc.3370010304>.
- [57] X. Zhang, W. Dahu, Application of artificial intelligence algorithms in image processing, *J. Vis. Commun. Image Represent.* 61 (2019) 42–49, <https://doi.org/10.1016/j.jvcir.2019.03.004>.
- [58] F.-Y. Gong, Z.-C. Zeng, F. Zhang, X. Li, E. Ng, L.K. Norford, Mapping sky, tree, and building view factors of street canyons in a high-density urban environment, *Build. Environ.* 134 (2018) 155–167, <https://doi.org/10.1016/j.buildenv.2018.02.042>.
- [59] X. Li, C. Ratti, I. Seiferling, Quantifying the shade provision of street trees in urban landscape: a case study in Boston, USA, using Google Street View, *Landsc. Urban Plann.* 169 (2018) 81–91, <https://doi.org/10.1016/j.landurbplan.2017.08.011>.
- [60] Z. Qavidel Fard, Z.S. Zomorodian, S.S. Korsavi, Application of machine learning in thermal comfort studies: a review of methods, performance and challenges, *Energy Build.* 256 (2022) 111771, <https://doi.org/10.1016/j.enbuild.2021.111771>.
- [61] L. Breiman, *Random Forests*, 45, 2001.
- [62] R. Yu, Y. Yang, L. Yang, G. Han, A. Oguti, RAQ-A random forest approach for predicting air quality in urban sensing systems, *Sensors* 16 (2016) 86, <https://doi.org/10.3390/s16010086>.
- [63] S. Chen, S.A. Billings, P.M. Grant, Non-linear system identification using neural networks, *Int. J. Control.* 51 (6) (1990) 1191–1214, <https://doi.org/10.1080/00207179008934126>.
- [64] J. Snoek, H. Larochelle, R.P. Adams, Practical bayesian optimization of machine learning algorithms, in: F. Pereira, C.J. Burges, L. Bottou, K.Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems*, 25, Curran Associates, Inc., 2012.
- [65] J. Kaldellis, D. Zafirakis, *Comprehensive Renewable Energy*, Elsevier, 2012.
- [66] H. Lan, H. Hou, Cynthia, Z. Gou, A machine learning led investigation to understand individual difference and the human-environment interactive effect on classroom thermal comfort, *Build. Environ.* 236 (2023) 110259, <https://doi.org/10.1016/j.buildenv.2023.110259>.
- [67] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, *arXiv* (2017), <https://doi.org/10.48550/ARXIV.1705.07874>.
- [68] N. Theeuwes, G.J. Steeneveld, R.J. Ronda, M. Rotach, B. Holtslag, Cool city mornings by urban heat, *Environ. Res. Lett.* 10 (2015) 4022, <https://doi.org/10.1088/1748-9326/10/11/114022>.
- [69] E. Sanagar Darbani, D. Monsefi Parapari, J. Boland, E. Sharifi, Impacts of urban form and urban heat island on the outdoor thermal comfort: a pilot study on Mashhad, *Int. J. Biometeorol.* 65 (2021) 1–17, <https://doi.org/10.1007/s00484-021-02091-3>.
- [70] F. Ali-Toudert, H. Mayer, Numerical study on the effects of aspect ratio and orientation of an urban street canyon on outdoor thermal comfort in hot and dry climate, *Build. Environ.* 41 (2) (2006) 94–108, <https://doi.org/10.1016/j.buildenv.2005.01.013>.
- [71] X. Yang, Y. Li, Z. Luo, P.W. Chan, The urban cool island phenomenon in a high-rise high-density city and its mechanisms, *Int. J. Climatol.* 37 (2) (2017) 890–904, <https://doi.org/10.1002/joc.4747>.
- [72] HKGBC, Sustainable building design guidelines (APP-152). <https://www.bd.gov.hk/doc/en/resources/codes-and-references/practice-notes-and-circular-letters/pmap/APP152.pdf>, 2016. (Accessed 22 January 2024).
- [73] SG, Singapore Green Plan 2030, Singapore Government, 2021. <https://www.greenplan.gov.sg/>. (Accessed 22 January 2024).
- [74] R.A. Johnson, D.W. Wichern, *Applied Multivariate Statistical Analysis*, fifth ed., Prentice Hall, 2002, pp. 354–425.