Renewable Energy 216 (2023) 119034

Contents lists available at ScienceDirect

Renewable Energy

journal homepage: www.elsevier.com/locate/renene

Fast and accurate estimation of solar irradiation on building rooftops in Hong Kong: A machine learning-based parameterization approach

Xuan Liao [a], Rui Zhu [b], Man Sing Wong [a,c,*], Joon Heo [d], P.W. Chan [e], Coco Yin Tung Kwok [a]

[a] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China
[b] Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, Singapore 138632, Republic of Singapore
[c] Research Institute for Land and Space, The Hong Kong Polytechnic University, Hong Kong, China
[d] Department of Civil and Environmental Engineering, Yonsei University, South Korea
[e] Hong Kong Observatory, Kowloon, Hong Kong, China

ARTICLE INFO

Keywords:
Rooftop solar irradiation
Machine learning
Parametric study
Geographic information system

ABSTRACT

Harvesting solar energy on rooftops can be a promising solution for providing affordable energy. This requires accurately estimating spatio-temporal solar photovoltaic (PV) potential on urban surfaces. However, it is still a challenge to obtain a fast and accurate estimation of rooftop solar PV potential over large urban built-up areas. Thus, this study proposes a parametric-based method to estimate annual rooftop solar irradiation at a fine spatial resolution. Specifically, seven parameters (Digital Surface Model, Sky View Factor, shadow from buildings, shadow from terrain, building volume to façade ratio, slope, and aspect) are determined that having great importance in modeling rooftop solar irradiation. Three machine learning methods (Random Forest (RF), Gradient Boost Regression Tree (GBRT), and AdaBoost) trained by the selected parameters are cross-compared based on R^2, Mean Absolute Error (MAE), and computation time. As a case study in Hong Kong, China, the RF outperformed GBRT and AdaBoost, with $R^2 = 0.77$ and $MAE = 22.83$ kWh/m^2/year. The time for training and prediction of rooftop solar irradiation is within 13 h, achieving a 99.32% reduction in time compared to the physical-based hemispherical viewshed algorithm. These results suggest that the proposed method can provide an accurate and fast estimation of rooftop solar irradiation for large datasets.

1. Introduction

1.1. Background

The global society is facing great challenges regarding increasing energy demand, climate change, and air pollution with accelerated global urbanization. Particularly, cities are responsible for 70% of overall global energy consumption, making up more than two-thirds of carbon emissions [1,2]. To mitigate these problems, governments need to implement effective policies to simultaneously curb the demand for fossil fuels and develop renewable energy [3]. Global solar photovoltaic (PV) power generation increased by 179 TWh in 2021 with a rough total generation of 1000 TWh, which makes solar PV become one of the major contributors to electricity generation in most of the countries [4,5]. Therefore, solar energy is a promising solution for satisfying the growing energy demands and for mitigating energy-related emissions in cities.

To realize the aim of Hong Kong's Climate Action Plan 2050, the Hong Kong Government has been promoting the utilization of renewable energy and has set an ambitious target on solar PV deployment in Hong Kong that all solar energy generation systems planned and approved in this Plan can generate about 200 million kWh of electricity each year, which is sufficient to meet the electricity demand of about 67,000 households [6]. Since Hong Kong is one of the densest cities with limited land to generate utility-scale green electricity, it is challenging to require large land space for concentrated solar farming. Rooftop solar PV system becomes a reasonable choice in terms of the availability of unused rooftop areas, easy installation and maintenance, and lower network transmission losses. However, optimal deployment of solar PV arrays on building rooftops still faces challenges in the accurate estimation of rooftop solar potential and the design of the intelligent distribution of PV deployment in existing studies.

* Correspondence to: Block Z, The Hong Kong Polytechnic University, 11 Yuk Choi Road, Hung Hom, Kowloon, Hong Kong.
E-mail address: Ls.charles@polyu.edu.hk (M.S. Wong).

https://doi.org/10.1016/j.renene.2023.119034
Received 20 March 2023; Received in revised form 11 June 2023; Accepted 12 July 2023
Available online 18 July 2023

1.2. Impact factors on rooftop solar irradiation

Spatio-temporal distribution of rooftop solar irradiation is affected by shade effects from buildings and mountains [7], rooftop slope and aspect [8], and a series of urban morphological features [9], such as the Sky View Factor (SVF). In the urban area, various artificial and natural objects create complex urban morphology, which increases shaded areas on building rooftops and decreases rooftop solar potential. Previous studies [10–14] have proven that urban morphological features influence building solar energy potential. Specifically, Zhu et al. [15] investigated the relationship between solar capacity and urban morphology features. The results illustrate urban morphology has an important effect on solar capacity influenced by the weather. Similarly, Poon et al. [16] conducted a parametric assessment to understand the correlation relationship between urban morphological features and annual average solar irradiation on rooftops and façades in Singapore. The results illustrate that SVF has the strongest correlation with solar irradiation on buildings. Furthermore, shade from surrounding buildings and mountains would directly and greatly reduce the amount of the received solar energy. Therefore, understanding the impact of these shades on rooftop solar irradiation is important for estimating and deploying PV arrays on building rooftops. Li et al. [17] indicated that the shade effects caused by building structures substantially influence the amount of installed power capacity and proposed a method to effectively compute the shaded areas on rooftops for accurately estimating the solar potential. Additionally, Digital Surface Model (DSM), rooftop slope, and aspect play vital roles in calculating solar irradiation in the model proposed by Rich et al. [18]. In conclusion, morphological features, shade effects, DSM, and rooftop slope and aspect are significant parameters for estimating rooftop solar irradiation. However, few existing studies investigate the specific impact of these parameters on evaluating rooftop solar potential. Our study not only analyzes the correlation of above-mentioned parameters with rooftop solar irradiation but also constructs the optimal machine learning model to explore the relationship between these parameters and solar irradiation.

1.3. Morphological tessellation

In the studies of urban morphology [19–21], the ratio between the footprint area and the unbuilt space is generally used to define density. And the calculation of density requires the definition of a boundary which is often based on a grid. Leng et al. [22] used a radius of 150 m as the investigation scale to calculate the urban morphological features (i.e., building site cover, floor area ratio, building height, road network density, road height-width ratio, green space ratio, and total wall surface area). Although the selection of the scale size is based on the empiricism of previous relevant studies [23–25], this selection method lacks reliable and accurate scientific supporting and the selected scale with regional limitations is difficult to be applied to other complex regions. To solve this problem, Yong et al. [26] investigated the spatial scale's influence on the accuracy of estimated results at spatial resolutions of 100, 200, 300, 400, 500, and 600 m. The results demonstrate that R square values grow as the spatial scale grows coarser, and this finding indicates that better prediction accuracy can be achieved using a coarse spatial scale. However, the approach of using a specific spatial scale for calculating density results in averaged values of a space portion, and it fails in capturing site-specific and building-related density morphological features. To overcome this limitation, Fleischmann et al. [27] presented Morphological Tessellation (MT) method to derive a spatial unit from the building footprint for urban morphometric analysis. And this study used MT method to generate morphological tessellation cells across four different urban tissues (i.e., organic tissue of Niederdorf, compact tissue of Langstrasse, detached villas of Hottingen, and mixed post-war development of Friesenberg) and had a visual inspection of these generated cells, and this suggests that MT method can be applied into other regions with similar urban tissues. Furthermore, Boccalatte et al. [28] used this MT method to calculate morphological features and evaluate the impact of urban morphology on rooftop solar radiation in Geneva. Since MT method can greatly evaluate the impact of each building on the surrounding space and accurately calculate building-related density information, this study employs MT method to calculate morphological features related to building density.

1.4. Methods for estimating rooftop solar potential

The estimation methods of roof solar potential can be classified into four methods, i.e., sampling method, geostatistical method, physical modeling method, and machine learning method [29]. The basic principle of the sampling approach is to calculate an estimate of available rooftop areas for the selected area, and then the estimate is extrapolated for the entire area. Specifically, Lzquierdo et al. [30,31] conducted stratified statistical sampling to compute the technical potential of rooftop photovoltaic (PV) energy production in Spain. The results showed that the total available rooftop area was approximately 571 km^2 across the country, and around 4% of the overall electrical energy can be generated by the PV system in Spain. Sampling methods can be used to extract the available rooftop area in a large-scale area, but these methods can only provide a rough estimation of rooftop solar potential, and they cannot satisfy the requirements of highly accurate estimation of rooftop solar irradiation.

The geostatistical methods performed spatial statistical analyses to predict solar potentials, such as spatial interpolation and statistical clustering-related methods. Fathizad et al. [32] proposed an air temperature-based model for the estimation of solar radiation and compared the evaluation performance of solar mapping using eight geostatistical methods, namely, Inverse Distance Weighted, Global Polynomial Interpolation, Radial Basis Function, Local Polynomial Interpolation, Ordinary Kriging, Simple Kriging, Universal Kriging, and Empirical Bayesian Kriging. The results suggested that the Radial Basis Function method is the most effective approach, with $R^2 = 0.904$, MAE = 3.02 $MJ/m^2/day$, and Root Mean Square Error (RMSE) = 0.39%. Additionally, Mishra et al. [33] used statistical clustering to calculate the available areas of rooftops for the estimation of rooftop solar potential in Uttarakhand, India. The results showed that 58% of rooftop area receives solar radiation greater than 4 $kWh/m^2/day$ throughout the year and could generate 57% of the electrical energy consumption in this region. Since geostatistical methods are usually focused on the total solar energy received, these methods can provide probabilistic estimations of solar potential. Sampling methods and geostatistical methods are limited in roughly and simply evaluating rooftop solar potential, so both methods face challenges in providing accurate and high spatio-temporal resolution estimates of the solar potential for an individual building.

GIS-based physical modeling methods have been considered as the optimal approach for estimating rooftop solar irradiation in terms of good estimation accuracy and the possibility of automated application in several areas. Saadaoui et al. [34] conducted an assessment method for solar PV potential on flat roofs in the city of Ben Guerir, Morocco using GIS and photogrammetry. The result indicated that more than 345 GWh of electricity can be annually generated by the rooftop solar PV system. Hong et al. [35] developed a method for estimating the hierarchical rooftop solar PV potential (i.e., physical, geographic, and technical potential) by analyzing the available rooftop area through Hillshade analysis in Seoul. The result illustrated that the physical, geographic, and technical potentials in the Gangnam district were found to be 9,287,982 MWh, 4,964,118 m^2, and 1,130,371 MWh, respectively. Although these methods can provide satisfactory results and investigate the multiple factors related to rooftop solar irradiation, it requires a much longer computation time for obtaining high-accuracy and reliable estimation results. For instance, Tabik et al. [36] proposed a Gradient

Ascent algorithm to compute the maximum irradiation based on GPU-CPU heterogeneous system, and it spent 2.477 s calculating a Digital Elevation Model (DEM) of 500 points. Therefore, these methods are more appropriately applied to micro-medium scale regions. Although the modeling-based methods for estimating solar potential have merits in terms of highly accurate estimates and a wide range of applications, these methods are limited to a large amount of execution time for large-scale regions.

In the last few years, machine learning methods have been increasingly used in studies related to the estimation of solar potential, with their merits on fast computation, application on a large scale, and providing high-accuracy and reliable results. Liao et al. [37] proposed a method to estimate spatially continuous land surface solar irradiation based on four machine learning methods in Australia and China, namely, Gradient Boosting Machine (GBM), Random Forest (RF), Support Vector Regression, and Multilayer Perceptron. The results showed GBM model achieved the highest accuracy with R^2>0.7 and the fastest computation ability using less than 10 s for processing hundreds of thousands of data. Likewise, Assouline et al. [38] combined solar models in GIS and the Random Forest algorithm to estimate the rooftop solar potential in 200×200 m^2 pixels in Switzerland. The results showed that the total estimated PV electricity production from building rooftops in Switzerland is 16.29 TWh/year, which can provide 25.3% of the yearly demand. Wang et al. [39] proposed a new PV power prediction model based on the Gradient Boost Decision Tree. The results suggested that this model has a good performance in model interpretation, prediction accuracy, and stable error performance. Babbar et al. [40] used Adaboost as a hybrid of linear and non-linear machine learning models for long-term solar power generation prediction. The hybrid Adaboost outperformed other individual machine learning models, with Percentage Mean Absolute Error (MAPE)= 8.88%. Compared to the aforementioned three methods, machine learning methods outperform them because of the fast computation applied in the large dataset and high accuracy for prediction. Therefore, it is reasonable to consider machine learning methods as appropriate alternative approaches for assessing and estimating rooftop solar potential.

1.5. Contributions

In summary, this study proposes a fast and accurate estimation method to estimate annual rooftop solar irradiation combined with multi-source data in Hong Kong, China. The main contributions of this study are: (i) performing the parametric assessment to understand the relationship between parameters related to the receiving rooftop solar irradiation and the corresponding annual solar irradiation on building rooftops, (ii) comparing different machine learning methods to estimate the annual rooftop solar irradiation at 1-m resolution in entire buildings in Hong Kong, and (iii) using a data and model dual-driven approach combined with the merits from machine learning methods and physical models.

This paper is organized as follows. Section 2 describes the data used in this study. Section 3 introduces the method for extracting morphological parameters and the methodology for estimating rooftop solar irradiation using machine learning models. Section 4 presents the estimated result and analyzes the distribution of rooftop solar irradiation in Hong Kong SAR. Section 5 offers conclusions and discusses the findings of the analysis.

2. Data

This section includes the description of the data sources, the data structure design adaptive to machine learning models, the study area, and data pre-processing.

2.1. Study area

Hong Kong, China, is located at 22° 15′ N, 114° 15′ E, with a typical subtropical climate. It has a total land area of about 1110 km^2 with a hilly and mountainous topography, and around 75% of the land in Hong Kong is a mountainous area. High population density and limited land resources have formed the high-density urban morphology in downtown areas of Hong Kong associated with densely packed high-rise buildings [22]. The territory is divided into 18 districts, with around 323,886 buildings by 2019. Because of the high building density and limited land resources, the rooftop solar PV system can be suitable and feasible for Hong Kong's renewable energy development in the future. Fig. 1 (b-g) shows the annual clear sky surface solar irradiance time evolution in six locations from 2012 to 2021 in Hong Kong, and the data is obtained from NASA Power [41]. To calculate the annual clear sky irradiance, we employed the monthly clear sky irradiance from NASA Power dataset. NASA power adopted the data from the Baseline Surface Radiation Network site observations. The monthly clear sky irradiance demonstrated a Bias of 0.03% and an RMSE of 5.7%. The resolution of the data is 1 degree latitude by 1 degree longitude. It is clear that the annual clear sky surface solar irradiance in different locations in Hong Kong is highly consistent, so this study does not use clear sky irradiance as the input variable. Fig. 1 (h-i) shows the amount of annual solar irradiation observed by the King's Park Station and Kau Sai Chau Station [42] from 2012 to 2021, which suggests that Hong Kong has a great potential for developing solar energy.

2.2. Dataset description

This study considers six influential factors that can affect spatio-temporal solar distribution on rooftops [7–9,17,43–45], namely morphological data, Digital Surface Model (DSM), building shadow, terrain shadow, tilted rooftop slope, and tilted rooftop aspect. Among them, the DSM at 1 m resolution and building polygons enriched with the height attribute were obtained from the Civil Engineering and Development Department and the Lands Department of the Government of Hong Kong SAR in 2019, respectively. This study intends to investigate the specific impact of building shade and mountain shade on rooftop solar irradiation, respectively. Therefore, this study calculates the building shade using the building footprint with building height and uses DEM to calculate the mountain shade. The rooftop solar irradiation map with 1 m resolution used for cross-validation is obtained from the project of Hong Kong Solar Irradiation Map for Building Rooftops which is conducted by the Electrical and Mechanical Services Department [46] and Remote sensing Lab at Hong Kong Polytechnic University, and it is calculated by using Remote Sensing technologies and Geographic Information Systems [47].

2.2.1. Urban morphological data

Previous studies have proved that urban morphology can affect the building solar energy potential [9,11,13–15,48–50]. Boccalatte et al. [28] evaluated the impact of 40 urban morphological parameters on rooftop solar radiation. Additionally, many studies [51–53] suggest that Sky View Factor (SVF) has a strong correlation with rooftop solar irradiation. Therefore, a total of 41 urban morphological parameters are calculated from building polygons in Hong Kong using a Python library named Momepy [54]. The Momepy library is based on several well-known Python packages for GIS-based data analysis, namely GeoPandas [55], PySAL [56], and networkX [57]. These 41 morphological parameters can be divided into four categories, and these can represent the building dimension, building shape, building intensity, and building spatial distribution. The list of these parameters, as well as the related equations and description, are displayed in Appendix A.

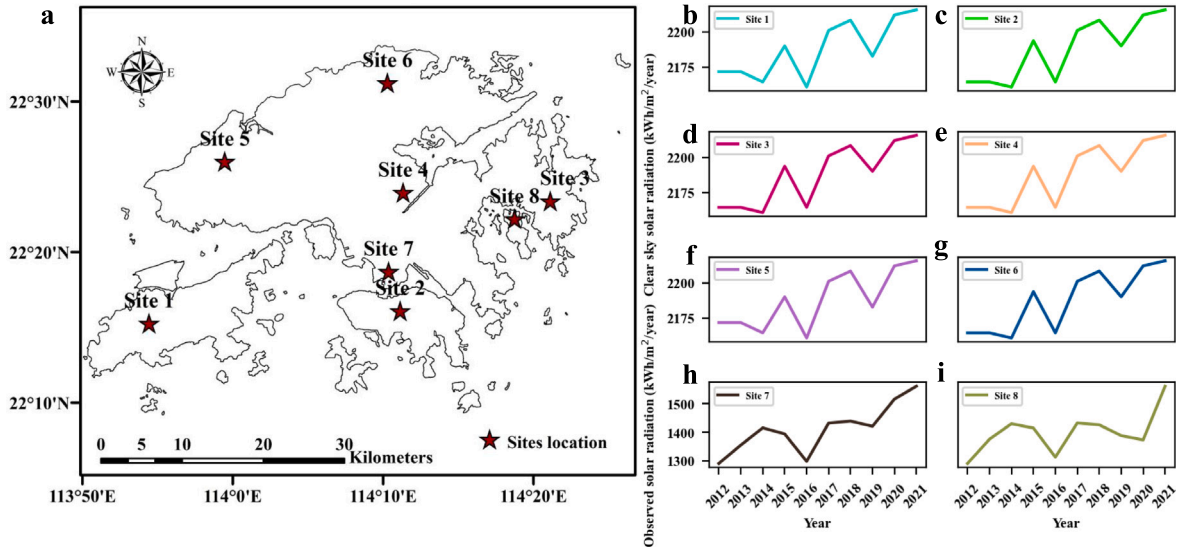

Fig. 1. The annual solar irradiance time evolution from 2012 to 2021 in Hong Kong, China. (a)Locations of eight sites. (b-g) Annual clear sky surface solar irradiance in six sites. (h-i) Annual solar irradiation from the King's Park Station and Kau Sai Chau Station.

2.2.2. Building shadow

Since skyscrapers in cities often cast shadows on each other [17], the mutual shadowing by buildings is considered in our study. A shadow polygon will be formed when the sunshine arrives at the rooftop. The direction of solar irradiation is described by the elevation and azimuth with a determined intensity at an instant of time.

2.2.3. Terrain shadow

Hong Kong is characterized by complex topography with high mountains and dense urban developments [43]. Therefore, the effect of terrain variation on rooftop solar irradiation is considered in this study.

2.2.4. Rooftop slope and aspect

Previous studies have suggested that tilted rooftops with various orientations significantly affect the site selection of solar PV arrays [7,8,44,45]. Thus, the rooftop characteristics, i.e., slope and aspect, are considered the input variables in our dataset. The *Aspect* and *Slope* toolsets in ArcMap generate the rooftop aspect image and slope image at 1-m resolution based on the DSM data.

2.3. Dataset pre-processing and data construction

Urban morphological data and rooftop slope and aspect are static data, which are directly calculated by Python Library and ArcMap. While building shadow and terrain shadow are dynamic data, both data need to be performed in accumulation processing for transforming into annual data. The building footprints and the height information are used to generate hourly 2D building shadow polygons from 7 am to 5 pm on 15th August 2019. Generated 2D building shadow polygons are transformed into Raster images, and these shadow images are overlayed into one day 2D building shadow image. The overlapped shadow image is considered daily shadow distribution for calculating the annual total building shadow image with a 1-m resolution. Figs. 2 (a) to (k) demonstrates hourly building shadow changes from 7 am to 5 pm, and Fig. 2(l) presents the accumulated shadow distribution on that day. Also, the *Hillshade* toolset in ArcMap is used to calculate hourly terrain shadow from 7 am to 5 pm on 15th August 2019 using the DEM data. Next, hourly shadow distributions are accumulated on a daily basis, and it is considered the average annual terrain shadow intensity at the resolution of 1 m.

All data are transformed into the raster files having the same resolution, orientation, and projection system (i.e., Hong Kong 1980 Grid coordinate system), which are deemed as multi-band images for training, validating, and testing. Our dataset contains 323,886 buildings that occupy 51-km^2 land surface in total. The whole data is organized according to the 18 districts. We divided these districts into the training and prediction regions (Fig. 3), which respectively account for 45% and 55% of the total rooftop area. This is based on two considerations. First, the training region covers high-density, middle-density, and low-density buildings, and the amount of building rooftops is well-sufficient for training and testing the models. Second, to evaluate the model performance, this study trains and validates the models and utilizes the resulting models to estimate the rooftop solar potential based on the testing dataset. The ratio of the training and testing datasets is 9 to 1. To improve the quality of the dataset, the outliers of all data values in datasets (i.e., the null value and infinite) which consumed 0.18% of the entire datasets are filtered out in the dataset.

3. Methodology

This study proposes a fast and accurate method based on the machine learning model for the estimation of annual rooftop solar irradiation over an urban area, with a flowchart presented in Fig. 4. Firstly, the Morphological Tessellation (MT) method [27] is used to calculate morphological features. Secondly, as a preliminary analysis to investigate the relationship between solar irradiation and the 41 morphological features A, Pearson correlation analysis has been performed to test the effectiveness of the proposed indices. To improve the training efficiency, Random Forest, a widely used machine learning model particularly useful for classification and prediction, is used to select suitable input variables. Furthermore, the estimation results are compared using different machine learning models and select the optimal model based on criteria of fast computation and the highest performance to estimate annual rooftop solar irradiation. Finally, this study analyzed the distribution of mean annual solar irradiation received by rooftops on different rooftop slopes and aspects for providing a reliable reference for effective deployment of solar PV arrays.

3.1. Calculation of morphological features

Forty-one morphological features used in this study are divided into four categories. The specific classification is shown in Appendix A. The features related to the categories of building dimension and building shape are directly calculated based on the building footprint. Additionally, this study employs the morphological tessellation cells (MTC) [27]

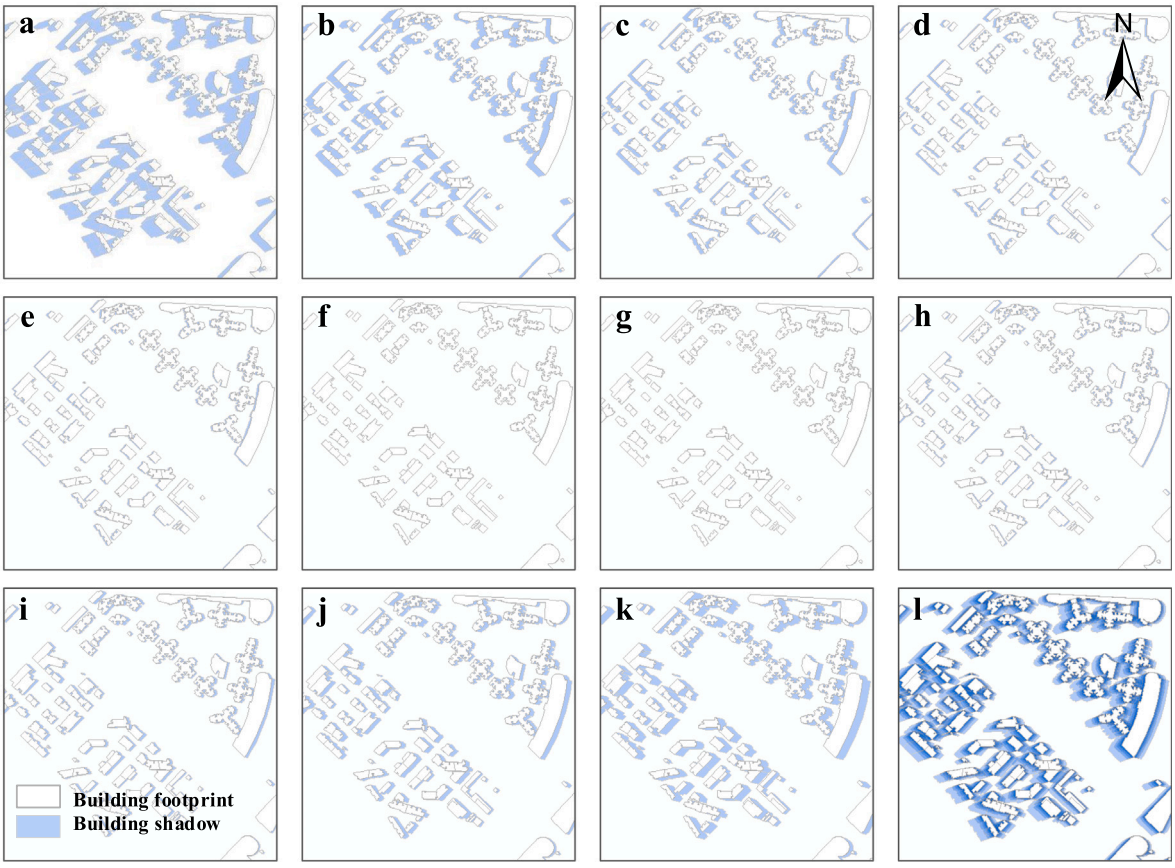

Fig. 2. Hourly shadow distribution in an urban area of Hong Kong on 15th August 2019. (a–k) Hourly shadow distribution from 7 am–5 pm. (l) Accumulated shadow distribution.

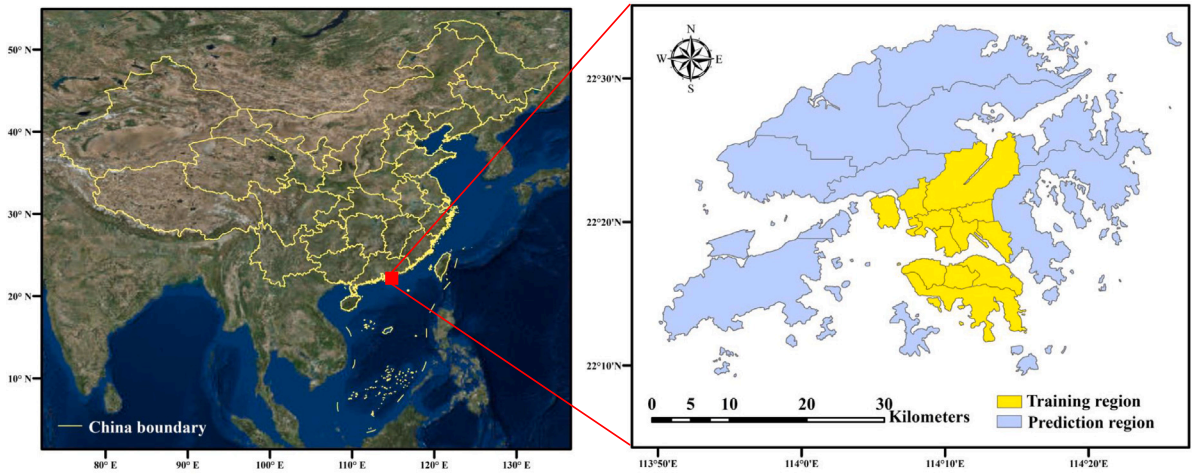

Fig. 3. The specific distribution of the training and testing regions.

to define a reference boundary for calculating the features related to building intensity. The aim of spatial distribution analysis is to calculate the spatial relationship among buildings, so the spatial distance of 200 m between a building and its adjacent buildings is employed based on the previous studies [58,59] for calculating the features related to building spatial distribution.

3.1.1. Morphological tessellation

Using a boundary for calculating the building density information requires the selection of specific spatial scale which is based on a grid or the administrative district for calculating building density [22]. However, this selection of an appropriate spatial scale usually relies on empiricism [23–26], and this leads to time-consuming for selecting the spatial scale and narrow scope of application. Compared to above-mentioned traditional method, the tessellation cell (MTC) is a geometric derivative of Voronoi polygons obtained from building footprints, and it represents the smallest spatial unit that delineates the portion of land around each building [28]. This allows us to obtain the density information related to buildings for a better estimation of solar distribution. The process of generating morphological tessellation consists of five steps: (i) inward offset from building footprint; (ii) discretization of polygons' boundaries into points; (iii) generation of Voronoi cells; (iv) dissolution of Voronoi cells; (v) clip of preliminary

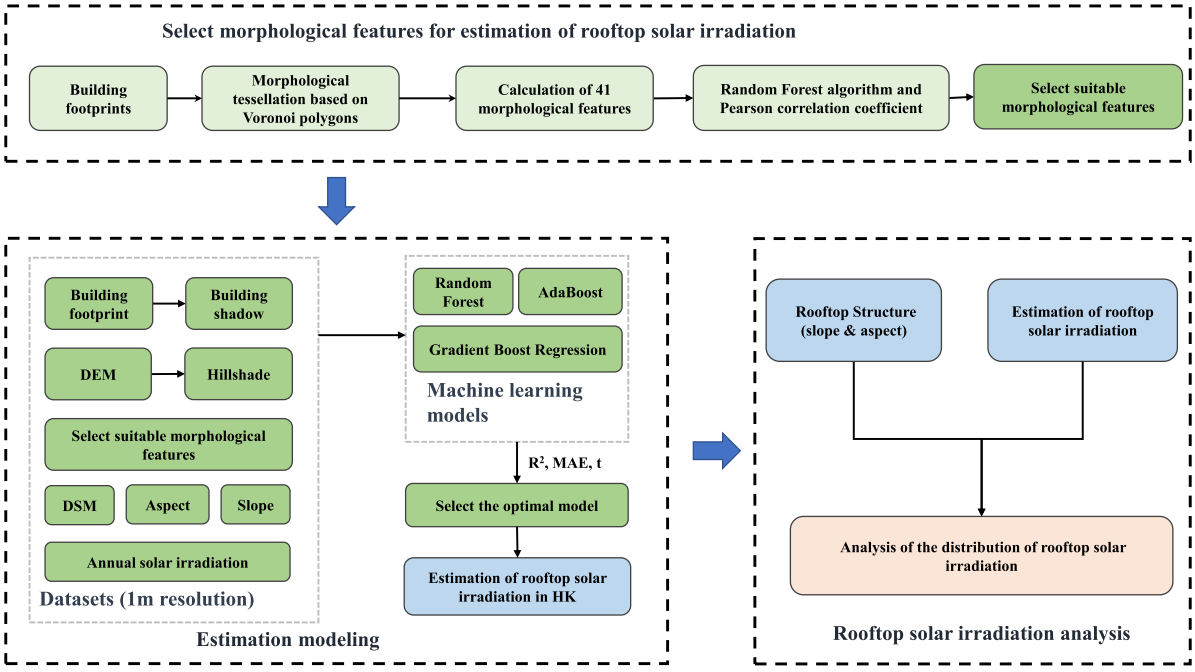

Fig. 4. The specific distribution of the training and testing regions.

tessellation. Fig. 5 shows the morphological tessellation distribution in Hong Kong. Based on the MTC, it is feasible to capture the influence that each building exerts on the surrounding space as well as the building-related density information.

3.2. Machine learning models

3.2.1. Random forest regression

RF [60] is a combined regression model that is composed of a large amount of decision tree $h(X;\theta_k)(k = 1,2,\ldots,n)$, where θ_k is an identically distributed random vector, n is the number of decision trees, and X is the input variables, namely, morphological indices, DSM, building shadow, terrain shadow, rooftop slope, and rooftop aspect. We set Y as the output variable that denotes the solar irradiation values calculated by the physical model, then (X,Y) composed of the original datasets. The RF model uses the Bootstrap method [61] for sampling of input datasets, and then employs the complete splitting method to construct the decision trees.

3.2.2. Gradient boost regression tree

GBRT [62] is an ensemble model using forward addition based on the base function of the classification and regression tree (CARF). The process of constructing one CARF consists of two parts, generation of the decision tree and decision tree pruning. In the process of constructing the GBRF model, the squared error is used as the learning target of the model (Eq. (1)):

$$L(y_i, f(x)) = min \sum_{i=1}^{N} (y_i - f(x))^2 \quad (1)$$

Where y represents the solar irradiation values calculated by the physical model as ground truth, x denotes seven input variables in our dataset, $f(x)$ represents the prediction value of the model, and N represents the size of the sample.

3.2.3. Adaboost regression tree

The Adaboost algorithm [63,64] is one of the best supervised learning methods with satisfactory prediction performance. This algorithm can inherit many weak regression models and finally form a strong model. The processing of the Adaboost algorithm is that for the same sample points, their weight will be continually updated for training multiple weak regression models, then the weak regression models with different weights are composed of a final strong regression model.

3.3. Selection importance parameters

The importance of each variable in RF can be estimated by the random sampling method. The original sample size is set as N, and variables are x_1, x_2,…, x_m (here $m = 46$, because we have 46 input variables in our dataset). Each time a bootstrap method is used to randomly select a sample from the total sample and randomly select n times with replacement. n bootstrap samples generate n regression trees. The samples that are not drawn each time are composed of n out-of-bag data as test samples, so that the importance order of each variable in the classification regression can be obtained.

3.4. Estimation model for annual rooftop solar irradiation

Physical modeling for the solar irradiation estimation are usually performed based on physical principles and mathematical modeling, thus these methods are often referred to as model-driven methods [65]. Model-driven methods have advantages of clear logistic and rigorous derivation. However, improving computing efficiency by simplifying model would always lead to decrease in the estimation accuracy [66]. Machine learning can extract knowledge from massive data with the advantages of high computing efficiency and high accuracy, thus these methods are known as data-driven methods. Nevertheless, these methods are dependent on a prior-knowledge and database. Since it is difficult to measure the ground truth of rooftop solar irradiation in large-scale regions, it faces challenges in obtaining ground truth for

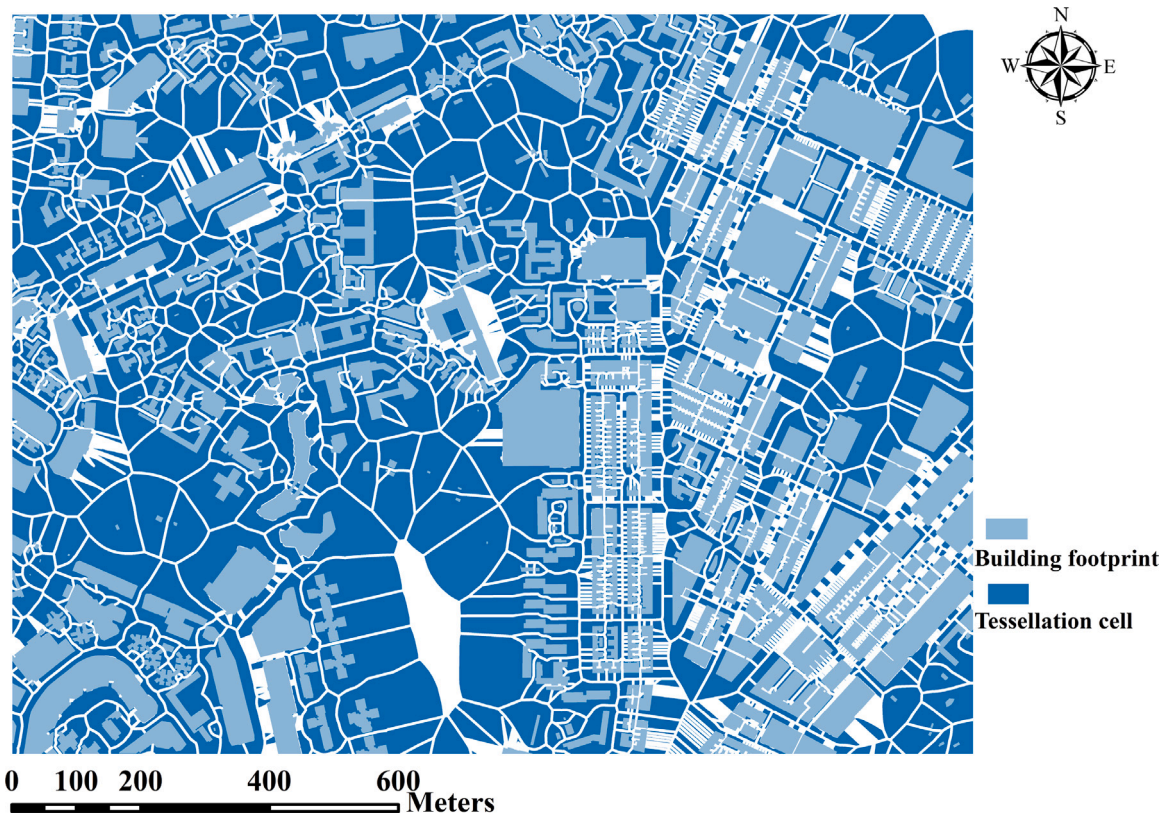

Fig. 5. Building footprints and related tessellation cells of a specific area of Hong Kong.

training the model. To rapidly and accurately estimate rooftop solar irradiation, a data and model dual-driven loosely coupled approach was proposed by the integrations of machine learning models and physical models. In this study, the estimation model consists of two modules, namely, the model-driven model and the data-driven model. The model-driven model uses the method proposed by Wong et al. [47], which is based on the hemispherical viewshed algorithm developed by Rich et al. [18], to calculate hourly rooftop solar irradiation. The specific hemispherical viewshed algorithm is introduced in Appendix B. The hourly solar data is accumulated into annual rooftop solar irradiation, and the model-driven model passes the annual data onto the data-driven model as the ground truth. In the data-driven model, three machine learning methods are compared to select the optimal one for estimating all rooftop solar potential by using three evaluating indicators, including R^2, Mean Absolute Error (MAE), and computation time.

4. Results

4.1. Correlation analysis between morphological features and rooftop solar irradiation

Pearson correlation analysis is performed to investigate the relationship between annual rooftop solar irradiation and 41 morphological features, and the Variance Inflation Factor (VIF) [67] is used to diagnose multicollinearity for these morphological features. These features can be divided into four categories, namely, building dimension, building shape, building intensity, and building spatial distribution [28]. Table 1 shows the Pearson correlation coefficients and corresponding p-values between rooftop solar irradiation and each morphological feature, and Table 2 shows multicollinearity among these features. The most explainable parameters relevant to rooftop solar irradiation are the ones related to the building shape (i.e., building shape index (Shp_{idx}), building Rectangularity (Rec), building equivalent rectangular index (ERI), building circular compactness (Com), building elongation (Elg), and building square compactness (Squ_{com})), with R>0.65. Furthermore, the parameters related to the building dimension (i.e., building fractal dimension (Fra), Tessellation longest axis length (LAL_{tess}), building volume to façade (VFR), and building longest exis length (LAL)) and ones related to the building spatial distribution, including, negative average neighborhood shading angle (Shd_{an}), building adjacency (Adj), sky view factor (SVF), and mean inter-building distance (IBD_{mean}). They show strong and positive correlations with R ranging between 0.65 and 0.7. On the contrary, the parameters with strong correlation coefficients do not consist of the ones related to the building intensity. For example, CAR has moderate correlations with rooftop solar irradiation, with R=0.61. The results indicate that the mentioned parameters related to the building shape, dimension, and spatial distribution can greatly affect the amount of the receiving rooftop solar irradiation. Only one feature, building shape index, has p-values between 0.01 and 0.005, which suggests that it is statistically significant at the 0.01 level. The p-values of four features (i.e., positive average neighborhood shading angle, average building height, rugosity, and alignment) are between 0.01 and 0.05, which suggests that they are statistically significant at 0.05 level. In addition, the p-values of the remaining 36 features are less than 0.005, which suggests that they are statistically significant at 0.005 level. The small values of VIF (VIF$\leq$ 10) corresponding to 11 morphological features suggest there is no issue with multicollinearity. In addition, the values of VIF corresponding to 16 morphological features are between 10 and 100, which suggests that these features have moderate multicollinearity, while the remaining morphological features whose values of VIF are larger than 100 suggest that they have strong multicollinearity.

4.2. Parameters selection and importance analysis

The RF model is often employed for calculating the parameter importance and selecting variables for training machine learning models

Table 1
Results of Pearson correlation analysis between rooftop solar irradiation and each morphological feature.

Group	Name	Name	R	Name	R
Fra	0.70***	CAR_{mean}	0.62***	A	0.47***
Shp_{idx}	0.70**	P	0.62***	HD_P	0.47***
Rec	0.69***	CAR	0.61***	V_{mean}	0.47***
ERI	0.69***	Shd_{ap}	0.61*	TFA_{mean}	0.46***
Com	0.68***	H	0.58***	Squ	0.42***
Adj	0.68***	H_{mean}	0.57*	SWR	0.42***
Shd_{an}	0.68***	FAR_{mean}	0.54***	V	0.40***
SVF	0.67***	Ort	0.53***	Flr_{area}	0.39***
Elg	0.67***	A_{mean}	0.51***	HW	0.35***
IBD_{mean}	0.66***	Rug	0.50*	H_{abs}	0.11***
Squ_{com}	0.66***	HD_n	0.49***	N_{neigh}	0.019***
VFR	0.65***	FAR	0.49***	HD	−0.28***
LAL_{tess}	0.65***	A_{tess}	0.47***	Shd_{mean}	−0.35***
LAL	0.65***	Ali	0.47*		

*p≤ 0.05.
**p≤ 0.01.
***p≤ 0.005.

Table 2
Results of multicollinearity analysis among morphological features.

Group	Name	VIF	Name	VIF	Name	VIF	Name	VIF
VIF >100	V_{mean}	8110	Rug	4542	Com	1069	LAL	109
	TFA_{mean}	7712	FA	4334	Fra	763		
	Flr_{area}	5393	Shp_{idx}	3463	Squ_{com}	469		
	V	5388	ERI	1485	Rec	181		
10<VIF≤ 100	P	96	FAR_{mean}	50	CAR	23	A_{mean}	17
	H_{mean}	54	CAR_{mean}	38	LAL_{tess}	22	Adj	14
	Shd_{an}	52	VFR	35	Shd_{ap}	21	N_{neigh}	14
	Elg	51	H	26	SVF	17		
VIF ≤ 10	IBD_{mean}	10	A_{tess}	6	Ort	3	HW	1
	HD	8	H_{abs}	6	A	3		
	HD_n	7	HD_P	5	Ali	2		
	SWR	7	Shd_{mean}	4	Squ	2		

in some studies [68–70]. The results are sorted in descending order as shown in Table 3. It presents that DSM makes a significant contribution to our estimation model, with 0.55 importance, followed by the rooftop slope. However, the importances of other parameters are close to zero. To increase the efficiency of the computation, we conducted recursive parameter selection for selecting useful parameters. In order to improve the efficiency of the selection, the interval of eliminating parameters was selected based on three indicators, i.e., R^2, MAE, and computation time. The order for eliminating parameters is based on the values of importance from low values to high values. Table 4 presents R^2, MAE, and computation time for recursively eliminating parameters from the parameter sets. The RF model with 46 parameters showed the highest estimation accuracy, with R^2=0.78. We further calculated the importance of these 46 parameters and arranged these parameters in descending order. The corresponding parameters are sequentially removed from the dataset parameter list, starting with the smallest value, based on their order of importance. Initially, the parameter removal interval was set at four. However, even with this interval, the R^2 value remained consistently at 0.78. As a result, the interval was adjusted to ten. However, this adjustment resulted in a slight increase in MAE. Consequently, the interval was further refined to eight. Considering the MAE and computation time, when the R^2 value decreased by 0.77, the interval was adjusted to one. Overall, as the number of parameters gradually decreases, corresponding R^2 and computation time also reduce, and MAE slightly increases. To balance the performance regarding the three indicators, models built by seven parameters are considered as appropriate for estimating rooftop solar irradiation, achieving high accuracy and fast computation. This is because the R^2 and MAE of the model with seven parameters are near

Table 3
The importance between rooftop solar irradiation and each parameter.

Name	I	Name	I	Name	I	Name	I
DSM	0.55	H_{mean}	0.0049	HW	0.0034	SWR	0.0027
Slope	0.13	Elg	0.0047	V	0.0034	FAR	0.0027
Shadow	0.055	IBD_{mean}	0.0045	Shd_{an}	0.0031	Rug	0.0025
Aspect	0.039	FAR_{mean}	0.0041	ERI	0.0031	TFA_{mean}	0.0024
SVF	0.026	LAL_{tess}	0.0040	HD_n	0.0030	Shp_{idx}	0.0024
Hillshade	0.021	Rec	0.0040	Flr_{area}	0.0030	V_{mean}	0.0023
VFR	0.016	CAR_{mean}	0.0039	HD	0.0030	Squ_{com}	0.0023
A_{tess}	0.0070	CAR	0.0038	Adj	0.0029	Com	0.0023
Ort	0.0064	HD_P	0.0038	Shd_{mean}	0.0028	H_{abs}	0.0023
Ali	0.0051	H	0.0038	Fra	0.0028	N_{neigh}	0.0021
Squ	0.0051	A_{mean}	0.0038	LAL	0.0027		
A	0.0051	Shd_{ap}	0.0037	P	0.0027		

Table 4
R^2, MAE, and time for recursively selecting parameters .

No. of input parameters	R^2	MAE (kWh/m^2/year)	Time (h)
46	0.78	20.71	20.47
42	0.78	20.72	19.27
38	0.78	20.73	17.10
28	0.78	20.77	12.50
20	0.78	20.94	8.54
14	0.78	21.21	6.01
7	0.77	22.83	3.00
6	0.74	24.77	2.50
5	0.74	24.63	2.10
4	0.70	27.33	1.70
3	0.67	27.74	1.30

to that of the model with 14 parameters, and using seven parameters can save half the computation time. The final dataset consists of DSM, shadow from the surrounding buildings, shadow from natural terrain, rooftop aspect, rooftop slope, building volume to façade ratio, and SVF.

4.3. Estimation of annual rooftop solar irradiation using machine learning models

The experiments were performed on a desktop with Intel Core i7-9700K CPU and 32 GB memory. Five-fold cross-validation [71] was performed to train and test each model. Specifically, the original dataset was randomly divided into five equally-sized sub-datasets. Among these five sub-datasets, one was designated as the validation data for evaluating the performance of machine learning models, while the remaining four sub-datasets were utilized as the training data. The grid search method [72] was used to optimize the hyper-parameters, and the optimization of hyper-parameters can be found in Table 5.

After selecting the final dataset, this study compared the estimation performance of three machine learning models in the Kowloon district (Table 6). The RF model using 46 parameters (RF_{46}) obtains the highest estimation accuracy with R^2 = 0.78 and MAE = 20.71 kWh/m^2/year. However, it costs 20.47 h to train a robust model. The value of R^2 of the RF model using seven parameters (RF_7) is close to that of the RF_{46} model, while the computation time of the RF_{46} model is around seven times longer. This suggests that the RF_{46} model is a complex estimation model with redundant parameters, which can lead to low computational efficiency. Although the AdaBoost model utilizing seven parameters ($AdBoost_7$) spends the least time for training, R^2 is only 0.58. This means that this model has a low capability to accurately estimate rooftop solar irradiation in the study area. The performance of RF_7 and GBM utilizing seven parameters (GBM_7) are followed by that of RF_{46}. The R^2 and MAE of the RF_7 datasets are higher than those of GBM_7, while the RF_7 model takes twice as long as the GBM_7 model.

To investigate the estimation accuracy of the models, this study also calculated the absolute errors. Table 7 shows the absolute error distribution in different ranges using four models in the Kowloon

Table 5
The hyper-parameters of the different machine learning models.

Model	The used hyper-parameters
RF_{46}	*n_estimators* : 200, *min_samples_leaf* : 1, *min_samples_split* : 2
RF_7	*n_estimators* : 100, *min_samples_leaf* : 1, *min_samples_split* : 2
GBM_7	*n_estimators* : 100, *learning_rate* : 0.1, *max_depth* : 3, *subsample* : 0.8
$AdBoost_7$	*n_estimators* : 50, *learning_rate* : 1.0, *base_estimator* : *deprecated*, *loss* : *linear*

Table 6
R^2, MAE, and time of different models in Kowloon.

Model	R^2	Mean absolute error (MAE) (kWh/m^2/year)	Time of training models (h)
RF_{46}	0.78	20.71	20.47
RF_7	0.77	22.83	3.00
GBM_7	0.71	28.72	1.47
$AdBoost_7$	0.58	42.25	0.87

Table 7
Absolute error distribution in different models in Kowloon.

Model	Range of the absolute error (kWh/m^2/year)			
	0–20	20–500	500–1000	>1000
$AdBoost_7$	92.53%	1.62%	5.86%	0.00%
GBM_7	92.67%	5.14%	1.87%	0.31%
RF_7	93.79%	4.39%	1.58%	0.24%
RF_{46}	94.15%	4.22%	1.36%	0.27%

district. The percentages of the absolute errors within 20 kWh/m^2/year are 94.15% for RF_{46}, 93.79% for RF_7, 92.67% for GBM_7, and 92.53% for $AdBoost_7$, respectively. For all models, the absolute errors over 500 kWh/m^2/year account for less than 6%. Compared with the four models, $AdBoost_7$ shows a slightly worse estimation performance, with around 8% of the absolute errors over 20 kWh/m^2/year. Overall, these four machine learning models show satisfactory estimation performance. This means that the estimation accuracy of all models is high and these models can provide reliable estimation results.

Combined with the results of Tables 6 and 7, the performance of estimation accuracy (i.e., R^2, MAE, and absolute error) of two RF models are better than the GBM model and AdaBoost model. To greatly investigate the computation efficiency of two RF models, this study compared calculation time for the calculation of the dataset, training model, and prediction using two RF models in Hong Kong. Table 8 is the result. Overall, the calculation time for each part using the RF_7 model is obviously less than those using the RF_{46} model. Especially, the training time using RF_{46} is more than 26 times longer than that of using RF_7. Thus, considering the estimation accuracy and computation time, we selected the RF model with seven parameters dataset for estimating the rooftop solar irradiation in the whole area of Hong Kong to obtain a balance between time cost and estimation accuracy.

The annual rooftop solar irradiation map was created by using the RF_7 model. Fig. 6(a) displays the annual rooftop solar map in Hong Kong, and Figs. 6 (b) to (e) show annual rooftop solar maps in Hong Kong Island, Central and West, Yuen Long, and Kowloon, respectively. The high density area shows smaller rooftop solar potential, while the low density area shows larger solar potential. This is because buildings in dense areas are greatly affected by the shadow from surrounding buildings. Therefore, shadow effect is a significant factor in estimating rooftop solar irradiation in dense cities.

To evaluate the usability and generalization ability of our model, this study compared Mean Relative Error (MRE) between the training and prediction regions (Tables 9 and 10). For training the model, the MRE of all the training regions is within 7%, and the time for training the model is approximately 12 h. For estimating the whole of Hong Kong, the MRE varies from about 9% to 5%, and the computation time is about 0.85 h. Although the MRE of the prediction regions is slightly higher than that in the training regions, this is a high estimation

Table 8
The comparison of calculation time for calculation of the dataset, training model, and prediction using two RF models in Hong Kong.

Model	Time for calculation of the dataset (h)	Time for training the model (h)	Time for prediction (h)
RF_{46}	32.79	319.79	2.41
RF_7	20.84	12.13	0.85

Table 9
The prediction accuracy in training regions.

Region	MRE	Region	MRE	Region	MRE
Central and Western	6.20%	Kwai Tsing	3.88%	Sham Shui Po	4.19%
East	4.57%	Kwun Tong	3.93%	Southern	4.50%
Yau Tsim Mong	4.58%	Sha Tin	4.73%	Wan Chai	5.69%
Wong Tai Sin	4.26%	Yau Tsim Mong	4.71%		

accuracy for the prediction regions which are not trained. The results indicate that our model has good generalization capability.

4.4. Accuracy assessment of physical model

The estimation of rooftop solar irradiation from the physical model was employed as the ground truth to cross-validate the machine learning models. To assess the accuracy of the physical model, field verification was conducted at five different sites, including a single-house rooftop in Kam Tin, a 20th-floor rooftop of private housing in Sha Tin, a sky garden at the Hong Kong Polytechnic University (PolyU), the lawn at the HKO King's Park Station, and a secondary school rooftop in Tseung Kwan O. Measurements were taken using MS-802, CM21, and CMP11 pyranometers. Table 11 shows the details of field verification, including data collection periods, site names and locations, and the equipment used. Additionally, Table 12 presents the comparison between validation field data and the estimated global horizontal solar irradiation using the physical model. Overall, the model achieves a high accuracy of 95.99% with an MRE of 4.01%. These results affirm the highly accurate performance of the physical model, validating the reliability of estimation values derived from it as ground truth.

4.5. Comparison of the physical model and machine learning model

The physical model demonstrated highly accurate results, with hourly estimations exhibiting 4.01% MRE for the entire year. Although the accuracy of the RF_7 model at 7.72% MRE is slightly lower than that of the physical model, both models can provide high accuracy and reliable estimation results. For comparison, the physical model and RF_7 model were used to estimate rooftop solar irradiation on 5334 buildings. These buildings cover 423,876 square meters randomly selected from the 18 districts in Hong Kong. The physical model spent around 927 s, versus the RF_7 model cost 6 s. From Table 8, it is clear that all calculation time required for estimating the whole rooftop solar potential using RF_7 model is approximately 33.82 h. In contrast, the physical model needs to spend nearly a year to complete the same estimation. This demonstrates that our model can greatly decrease computation time and thus overcome the low-efficiency problem when using the physical model. For the input parameters, the physical model utilized DSM, locational information (latitude and longitude), slope, aspect, and direct and diffuse solar radiation to calculate the rooftop

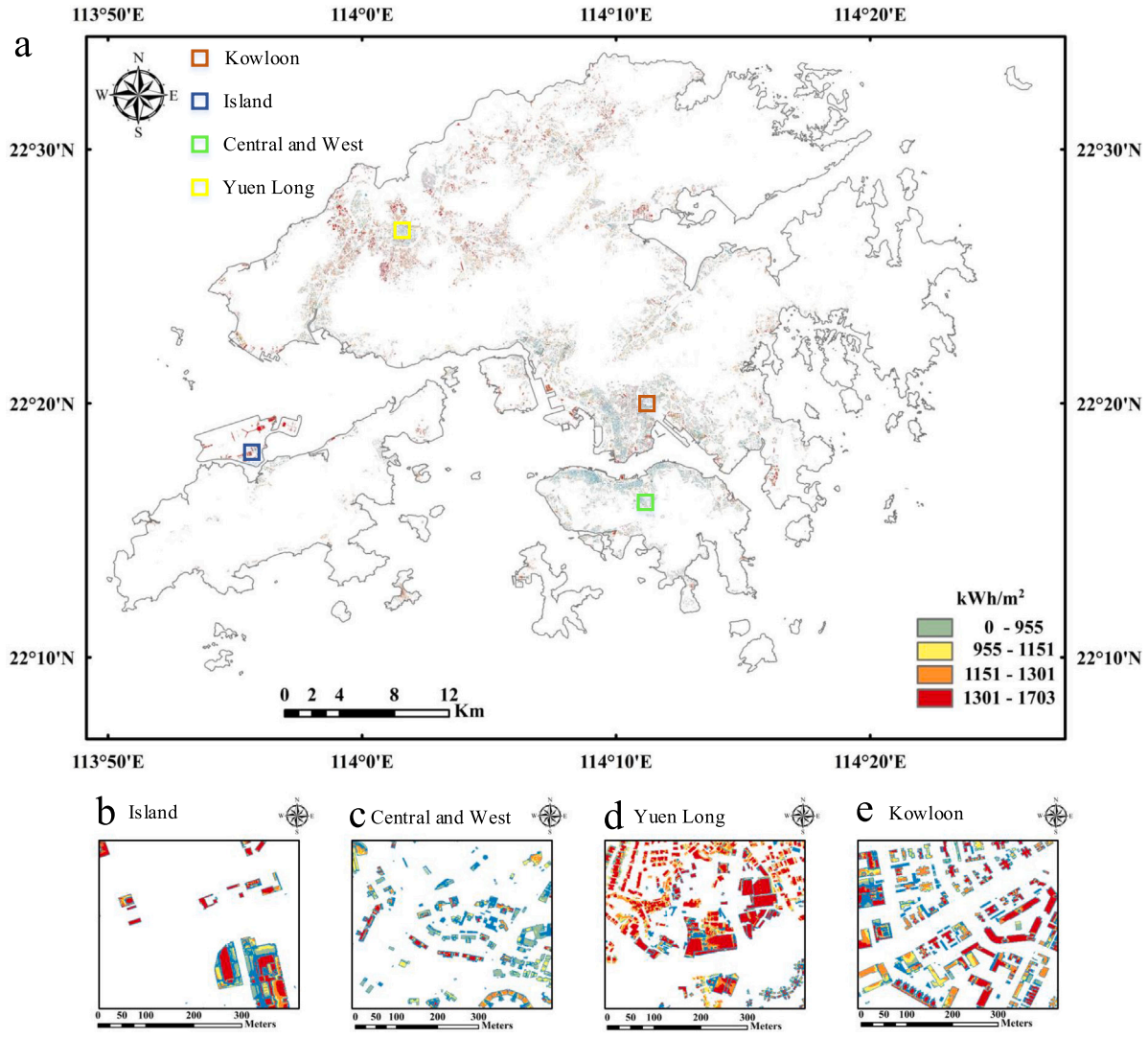

Fig. 6. Annual rooftop solar map in Hong Kong. (a) The whole solar map. (b) specific solar map in Hong Kong Island. (c) specific solar map in Central and West. (d) specific solar map in Yuen Long. (e) specific solar map in Kowloon.

Table 10
The prediction accuracy in prediction regions.

Region	MRE	Region	MRE	Region	MRE	Region	MRE
Hong Kong Island	9.11%	Sai Kung	7.17%	Tai Po	8.37%	Tsuen Wan	9.27%
Tuen Mun	9.30%	Yuen Long	9.59%	North	5.16%		

Table 11
Details of field verification.

Site	Period	Location Name	Coordinates	Equipment Used
1	22 Feb 2020–25 Feb 2020	Kam Tin	(22.24, 114.07)	MS-802, CM21
2	25 Feb 2020–28 Feb 2020	Sha Tin	(22.38, 114.20)	MS-802, CM21
3	29 Apr 2020–6 May 2020	PolyU	(22.31, 114.18)	MS-802, CM21
4	27 Aug 2020–7 Sep 2020	King's Park Station	(22.31, 114.17)	MS-802, CM21, CMP11
5	30 Dec 2020-6 Jan 2021	Tseung Kwan O	(22.32, 114.26)	CMP11

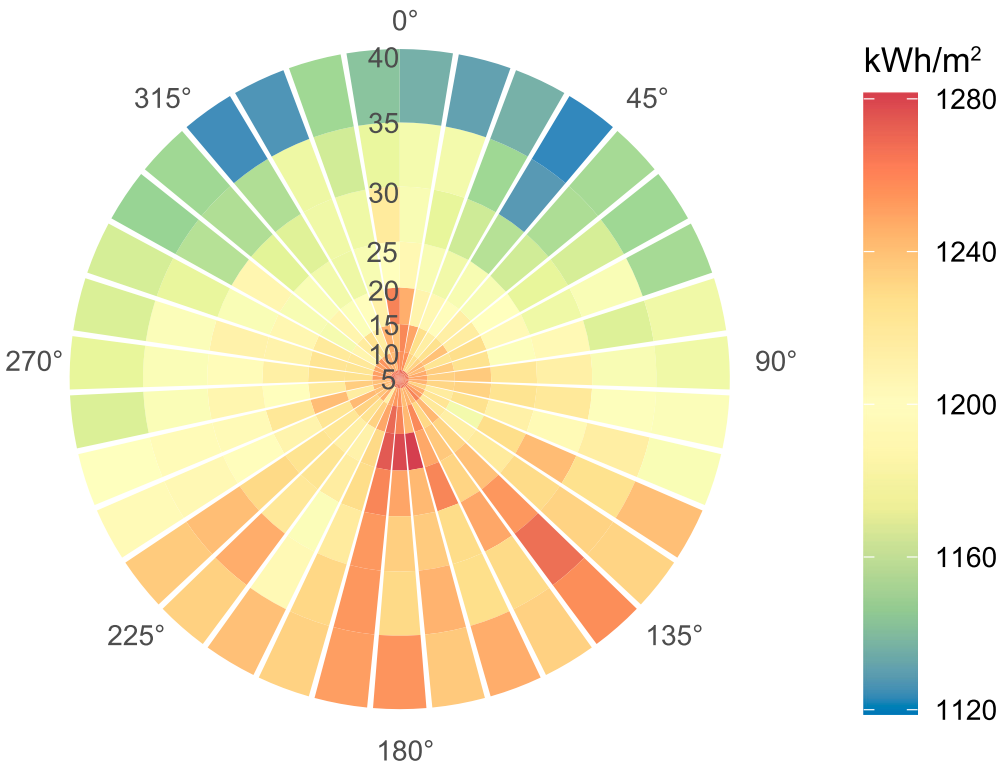

Fig. 7. Annual mean solar radiation (kWh/m^2) as a function of slope and aspect of roof surfaces for buildings in Hong Kong, China.

Table 12
Comparison between validation field data with the estimated result at the five validation sites.

Site	Estimated result (Wh/m^2)	Measurement (Wh/m^2) (MRE)		
		MS-802	CM21	CMP11
1	18,979	19,092 (−0.59%)	19,711 (−3.71%)	N/A
2	12,771	11,331 (12.71%)	11,408 (11.95%)	N/A
3	35,914	35,964 (−0.14%)	37,508 (−4.25%)	N/A
4	42,631	43,038 (−0.94%)	44,412 (−4.01%)	44,264 (−3.69%)
5	24,357	24,357	24,357	24,232 (0.51%)

solar irradiation, obtaining high spatial–temporal resolution of solar radiation. Compared to the physical model, the RF_7 model only uses seven parameters (i.e., DSM, shadow from the surrounding buildings, shadow from natural terrain, rooftop aspect, rooftop slope, and SVF) to get highly accurate and efficient estimation, and these parameters are relatively easy to obtain a reliable generalization.

4.6. Analysis of rooftop solar irradiation distribution

After the estimation of annual mean solar irradiance by applying our trained model, we conducted an analysis to explore the relationship between annual mean solar irradiance and slopes as well as aspects. This analysis involved associating the estimated solar irradiance values with their corresponding slopes and aspects, which were determined based on the geographical locations of the rooftops. Fig. 7 visualizes the average annual solar irradiation received by the rooftop as a function of roof slope and aspect. Overall, the annual mean solar irradiation received by rooftops is high, from 1120 kWh/m^2 to 1280 kWh/m^2. This suggests that solar potential on rooftops in Hong Kong could generate a considerable amount of electricity efficiently. The distribution of rooftop solar irradiation is east–west symmetry, and the values of solar irradiation gradually decrease from north to south. Furthermore, the largest irradiation is found for south-facing rooftops with a slope between 30 and 40 degrees, while north-facing rooftops have the lowest solar irradiation, smaller than 1200 kWh/m^2. The results are in line with the order of nature that the sun shines mainly from the south to the north in the northern hemisphere.

The team also investigated annual mean solar irradiation on rooftops surfacing different aspects and slopes, respectively. Results in Fig. 8(a) show that the rooftops facing south receive the greatest solar irradiation, followed by rooftops facing west and east; in particular, the north receives the least irradiation, as expected. Fig. 8(b) shows that the flat to gently sloping rooftops, with the slope ranging from 0–40 degrees, receive the greatest solar irradiation. The results illustrate the steeper the slope, the smaller the received solar irradiation.

5. Discussion and conclusion

Obtaining a fast and accurate estimation of rooftop solar PV potential over large urban built-up areas poses several challenges, including processing vast amounts of data, selection of appropriate impact factors, as well as appropriate models. To overcome these limitations, this study proposes a data and model dual-driven loosely coupled method. This innovative method enables fast and accurate estimation of annual rooftop solar irradiation at a spatial resolution of 1 m in Hong Kong, China. This study parameterizes the influential factors (i.e., morphological features, building rooftop structures, DSM, the shadow from buildings, and the shadow from terrain) and quantifies the importance of these features on rooftop solar irradiation. Compared between RF, GBM, and AdaBoost, the RF model is used for the estimation of annual rooftop solar irradiation for individual buildings since its estimation

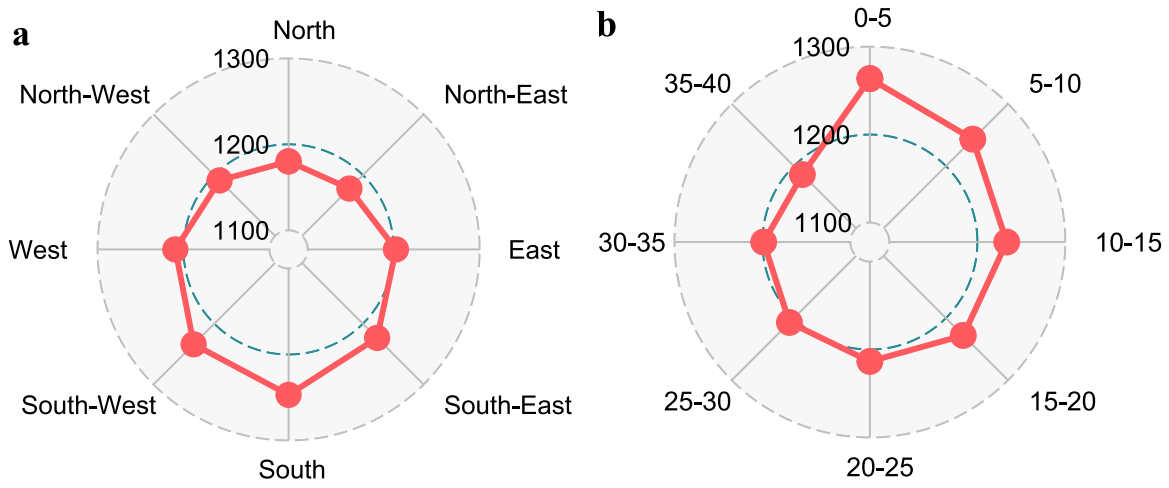

Fig. 8. Annual mean solar irradiation (kWh/m^2) of roof surfaces for different ranges of (a) aspect and (b) slope.

accuracy is high (R^2 = 0.77), and the computation speed is fast. Compared with the physical model, the machine learning models developed in this study can greatly reduce the computation time for rooftop solar estimation at fine spatio-temporal scales. These results suggest that our method can estimate rooftop solar irradiation on individual buildings, which is useful for solar related applications, such as planning rooftop PV arrays. As our developed models are well-trained and validated with a satisfactory scalability in various spatial and temporal resolutions, it is possible to apply these models to other regions having a similar built environment, and the proposed method is also deliverable for entirely different areas.

The traditional methods of calculating building density usually require the definition of a reference boundary, which is generated by a grid or administrative limits of a district. However, these methods just calculate the average value in a certain portion and fail in capturing site-specific and density information related to buildings. The morphological tessellation method used in our study can overcome this limitation, which makes it possible to capture the specific impact of surrounding space on each building.

This study approximated the estimated rooftop solar irradiation from the physical model as the ground truth for cross-validation. This is because getting field measurements by installing high density of solar sensors on all the rooftops in Hong Kong is almost impossible. This is one of the feasible solutions as previous studies also utilized the similar method for validation [63,73].

Compared with the conventional physical models, such as the upward-looking hemispherical viewshed algorithm, our approach is 5592 times faster in computing annual solar potential on all rooftops in Hong Kong. Therefore, we conduct recursive parameters selection to filter out redundant parameters based on the balance of estimation accuracy and computation time. Results of the model with seven parameters show high accuracy with fast computation, and this indicates that this model enables to satisfy the requirements for estimating rooftop solar irradiation in terms of accuracy and computation speed.

However the developed method outperformed compared with others, it has some limitations, i.e. we calculated building shadow and terrain shadow on one specific day to represent annual shadow distribution, which would affect the estimation accuracy in some extent. This is because the calculation of hourly building shadow and terrain shadow for one day with high spatio-temporal resolution requires a computation time around 24 h. In this regard, this study uses this estimation method for shadow data. From the final results of the estimation of rooftop solar irradiation, the method proposed in this study can provide high estimation accuracy. Therefore, using a generalized shadow data can decrease the computation time and confirm estimation accuracy at the same time.

In conclusion, we propose a fast and accurate parametric method for estimating rooftop solar irradiation based on the machine learning method using seven parameters (DSM, SVF, shadow from buildings, shadow from mountains, VRF, slope, and aspect). The results demonstrate that the proposed method can provide a reliable, fast, highly accurate reference for potential applications, including solar PV installation planning, financial analysis and investment decision-making, and urban planning. Specifically, our results can help relevant parties to identify suitable locations for solar PV installations and make decisions on the feasibility and optimal placement of solar panels on rooftops. The method also can help to assess the potential solar energy generation and associated cost savings, enabling governments and companies to evaluate the viability and profitability of rooftop solar PV installations. Additionally, the highly accurate estimation results can provide reliable references to optimize the orientation and layout of future constructions and maximize solar energy utilization which can help to effectively reduce the Urban Heat Island.

CRediT authorship contribution statement

Xuan Liao: Conceptualization, Methodology, Investigation, Validation, Visualization, Writing – original draft, Writing – review & editing. **Rui Zhu:** Methodology, Visualization, Writing – original draft, Writing – review & editing. **Man Sing Wong:** Conceptualization, Methodology, Investigation, Supervision, Writing – original draft, Writing – review & editing. **Joon Heo:** Writing – original draft, Supervision. **P.W. Chan:** Supervision. **Coco Yin Tung Kwok:** Validation, Writing – original draft.

Declaration of competing interest

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

Acknowledgments

This project is supported from the General Research Fund [Grant No. 15602619] from the Research Grants Council, Hong Kong, China. M.S. Wong also thanks the funding support from the General Research Fund [Grant No. 15603920], the Collaborative Research Fund [Grant No. C5062-21GF] from the Research Grants Council, Hong Kong, China, and project [1-CD81] from the Research Institute for Land and Space, the Hong Kong Polytechnic University, Hong Kong, China.

Appendix A

See Table A.1.

Table A.1
Morphological features.

Name	Description	Symbol	Category	Equation
Building height	Building height	H	D	–
Building area	Building footprint area	A	D	–
Building volume	Building volume	V	D	–
Building perimeter	Sum of lengths of the building exterior walls	P	D	–
Building longest axis length	Diameter of the minimal circumscribed circle around the building footprint	LAL	D	–
Building volume to façade ratio	Ratio between building volume and the total area of façades	VFR	D	$\frac{volume}{perimeter \bullet height}$
Building fractal dimension	Statistical index of the complexity of a geometry	Fra	D	$\frac{2\log(perimeter/2)}{\log(area)}$
Building circular compactness	Index of the similarity of a shape with a circle. It is based on the area of the minimal enclosing circle (Ac)	Com	S	$\frac{area}{Ac}$
Building square compactness	Measure of the compactness of the building footprint	Squ_{com}	S	$\left(\frac{4\sqrt{area}}{perimeter}\right)^2$
Building squareness	Mean deviation μ of each i corner of the building from 90°. Ncor is the number of corners	Squ	S	$\frac{\sum_{i=1}^{Ncor}\mu_i}{Ncor}$
Building Rectangularity	Index of the similarity of a shape with a rectangle. It is based on the area of the minimal rotated bounding rectangle of the building (AMBR)	Rec	S	$\frac{area}{AMBR}$
Building shape index	Shape index of the building footprint	Shp_{idx}	S	$\frac{\sqrt{\frac{area}{\pi}}}{0.5 \bullet lal}$
Building equivalent rectangular index	Measure of shape complexity based on the area of the minimal rotated bounding rectangle of a building (AMBR) and its perimeter (PMBR)	ERI	S	$\sqrt{\frac{area}{AMBR}} \bullet \frac{PMBR}{perimeter}$
Building elongation	Measure of the deviation of the building shape from a square based on the length of the minimal rotated bounding rectangle of a building (LMBR) and its width (lMBR)	Elg	S	$\frac{LMBR}{IMBR}$
Floor area ratio	Ratio between the building total floor area and the area of the related tessellation cell	Flr_{area}	I	$\frac{area}{t_{area}}$
Shared walls ratio of adjacent buildings	Ratio between the length of the perimeter shared with adjacent buildings (Pshared) and the building perimeter	SWR	SD	$\frac{Pshared}{perimeter}$
Building orientation	Building orientation	Ort	D	–
Alignment	Mean deviation of solar orientation (devsol) of neighboring buildings	Ali	SD	$\frac{\sum_{j\in neigh} devsol(j)}{Nneigh}$
Building adjacency	Ratio between the number of joined adjacent structures (Nneigh,join) and the number of neighboring buildings (Nneigh)	Adj	SD	$\frac{Nneigh_{join}}{Nneigh}$
Mean inter-building distance	Mean distance between the building and the adjacent buildings	IBD_{mean}	SD	$\frac{1}{Nneigh}\sum_{j\in neigh} d(j)$
Average building area	Mean footprint area of building neighboring constructions	A_{area}	SD	$\frac{1}{Nneigh}\sum_{j\in neigh} area(j)$
Average building height	Mean height of building neighboring constructions	H_{mean}	SD	$\frac{1}{Nneigh}\sum_{j\in neigh} height(j)$
Average building volume	Mean volume of building neighboring constructions	V_{mean}	SD	$\frac{1}{Nneigh}\sum_{j\in neigh} volume(j)$
Average building total floor area	Mean total floor area of building neighboring constructions	TFA_{mean}	SD	$\frac{1}{Nneigh}\sum_{j\in neigh} floor_{area}(j)$
Average Height to Width ratio	Mean ratio between building height and width of building neighboring constructions	HW	SD	$\frac{1}{Nneigh}\sum_{j\in neigh} \frac{H}{d(j)}$
Distance-weighted average height difference	Mean height difference with distance weighted between the reference building and its neighboring buildings	HD	SD	$\frac{\sum_{j\in neigh}(H(j)-H)\bullet w(j)}{\sum_{j\in neigh} w(j)}$
Average neighborhood shading angle	Mean shading angle between the reference building and its neighboring buildings	Shd_{mean}	SD	$arctan\left(\frac{1}{Nneigh}\sum_{j\in neigh}\frac{H(j)-H}{d(j)}\right)$
Positive distance-weighted average height difference	Mean height difference with distance weighted between the reference building and its neighboring buildings (H(j)>H)	HD_p	SD	$\frac{\sum_{j\in neigh}(H(j)-H)\bullet w(j)}{\sum_{j\in neigh} w(j)}$
Negative distance-weighted average height difference	Mean height difference with distance weighted between the reference building and its neighboring buildings (H(j)<H)	HD_n	SD	$\frac{\sum_{j\in neigh}(H(j)-H)\bullet w(j)}{\sum_{j\in neigh} w(j)}$
Positive average neighborhood shading angle	Mean shading angle between the reference building and its neighboring buildings (H(j)>H)	Shd_{ap}	SD	$arctan\left(\frac{1}{Nneigh}\sum_{j\in neigh}\frac{H(j)-H}{d(j)}\right)$
Negative average neighborhood shading angle	Mean shading angle between the reference building and its neighboring buildings (H(j)<H)	Shd_{an}	SD	$arctan\left(\frac{1}{Nneigh}\sum_{j\in neigh}\frac{H(j)-H}{d(j)}\right)$
Rugosity	Ratio between the building volume and the area of the related tessellation cell	Rug	I	$\frac{volume}{area_t}$
Floor area	Floor area of each object based on height and area	FA	S	$\frac{height \bullet area}{3}$
Coverage area ratio	Ratio between the building footprint area and the area of the related tessellation cell	CAR	I	$\frac{area}{area_t}$
Mean coverage area ratio	Mean coverage area ratio of the neighboring tessellation cells	CAR_{mean}	SD	$\frac{1}{Nneigh}\sum_{j\in neigh} CAR(j)$
Mean floor area ratio	Mean floor area ratio of the neighboring tessellation cells	FAR_{mean}	SD	$\frac{1}{Nneigh}\sum_{j\in neigh} FAR(j)$
Sky view factor	Sky view factor	SVF	SD	–
Number of neighbors	Number of neighbors	N_{neigh}	SD	–
Tessellation longest axis length	Diameter of the minimal circumscribed circle around the tessellation cell	LAL_{tess}	D	–
Average tessellation area	Mean tessellation area of building neighboring tessellation cells	A_{tess}	SD	$\frac{1}{Nneigh}\sum_{j\in neigh} area_{t(j)}$

Appendix B. Upward-looking hemispherical viewshed algorithm

The estimates of rooftop solar irradiation calculated by the Solar Analyst Tool in ArcMap are as the ground truth in our machine learning model training. The Solar Analyst Tool is based on methods from the hemispherical viewshed algorithm developed by Fu and Rich [18]. The specific process of the upward-looking hemispherical viewshed algorithm is as follows.

B.1. Viewshed calculation

A viewshed is the angular distribution of sky visibility versus obstruction and represents the proportion of the obstructed sky in a specific location on a DEM, which is similar to the view from upward-looking hemispherical photographs. Viewsheds are calculated by searching a specified set of directions around an interesting location on DEM in each direction and determining the maximum angle of sky obstruction. For the unsearched directions, the interpolation method is used to calculate the horizon angles. The horizon angles are projected into a two-dimensional (2D) grid using an equiangular hemispherical projection. A value with visible versus obstructed sky directions is assigned to each corresponding grid unit. The grid cell location (i.e., row and column) represents a zenith angle θ and an azimuth angle α on the hemisphere of directions.

B.2. Sunmap calculation

After generating a viewshed for a specific location on a DEM, a sunmap is created to represent the amount of direct solar radiation from each sky direction in the same 2D grid system. The sunmap consists of specified suntracks, and it represents the apparent position of the sun as it varies through time. Zenith and azimuth angles are used for representing the position of the sun, and they are calculated based on latitude, day of year, and time of day using standard astronomical formulae [18]. Zenith and azimuth angles are projected to 2D grids with the same resolution as the viewsheds. Two sunmaps are created, namely, sunmap for winter solstice to summer solstice, and sunmap for summer solstice to the winter solstice. For each sky sector of the sunmap, the associated time duration and the position of the sun are calculated, and each sector is assigned a unique identification number.

B.3. Skymap calculation

To achieve the skymap calculation, the whole sky is divided into a series of sky sectors defined by zenith and azimuth divisions. The skymap is used in the final solar radiation calculation to estimate diffuse solar radiation. The sky sectors in the skymap are required to be small enough that the centroid zenith and azimuth angles can reasonably represent the direction of the sky sector in subsequent calculations. The skymap is also projected into the 2D grid for the final solar radiation calculation.

B.4. Overlay of viewsheds with sunmaps and skymaps

After creating sunmap and skymap, two maps are overlayed to enable calculation of the direct and diffuse solar radiation received from each sky direction. For gap fraction in the skymap or sunmap sector, it is calculated by dividing the number of unobstructed units by the total of units in that sector.

B.5. Global solar radiation calculation

Since reflection radiation accounts of small proportion of the global solar radiation, global radiation G_R is calculated as the sum of direct and diffuse radiation of all sectors in the sunmap and skymap. The formula is as follows:

$$G_R = D_R + F_R \tag{B.1}$$

where D_R denotes the total direct solar radiation for all sunmap sectors, F_R represents the total diffuse solar radiation for all skymap sectors. The formula is as follows:

$$D_R = \sum D_{\theta,\ \alpha} \tag{B.2}$$

where $D_{\theta,\ \alpha}$ denotes the direct insolation from the sunmap sector with a centroid at zenith angle θ and azimuth angle α. The formula for calculation of $D_{\theta,\ \alpha}$ is as follows:

$$D_{\theta,\ \ \alpha} = S_{const} \times \tau^{m(\theta)} \times SunDur_{\theta,\ \alpha} \times SunGap_{\theta,\ \alpha} \times \cos\left(AngIn1_{\theta,\ \alpha}\right) \tag{B.3}$$

where S_{const} denotes a solar constant and the its range is from 1338 to 1368 WM^{-2}, τ is transmittivity of the atmosphere for the shortest path, m(θ) is the relative optical path length, $SunDur_{\theta,\ \alpha}$ the time duration represented by the sky sector, $SunGap_{\theta,\ \alpha}$ is the gap fraction for the sunmap sector, and $AngIn1_{\theta,\ \alpha}$ is the angle of incidence between the centroid of the sky sector and the axis normal to the surface. The formula for calculation of F_R is as follows:

$$F_R = R_{glb} \times P_{dif} \times Dur \times SkyGap_{\theta,\alpha} \times Weight_{\theta,\alpha} \times \cos\left(AngIn2_{\theta,\ \alpha}\right) \tag{B.4}$$

where R_{glb} is the global normal radiation, P_{dif} is the proportion of global normal radiation flux that is diffused, Dur is the time interval for analysis, $SkyGap_{\theta,\alpha}$ is the gap fraction (proportion of visible sky) for the sky sector, $Weight_{\theta,\alpha}$ is proportion of diffuse radiation originating in a given sky sector relative to all sectors. $AngIn2_{\theta,\ \alpha}$ is the angle of incidence between the centroid of the sky sector and the intercepting surface.

References

[1] A. Alhamwi, W. Medjroubi, T. Vogt, C. Agert, Development of a GIS-based platform for the allocation and optimisation of distributed storage in urban energy systems, Appl. Energy 251 (2019) 113360.

[2] F. Jalil-Vega, I. Kerdan, A. Hawkes, Spatially-resolved urban energy systems model to study decarbonisation pathways for energy services in cities, Appl. Energy 262 (2020) 114445.

[3] DM. Kammen, DA. Sunter, City-integrated renewable energy for urban sustainability, Science 352 (2016) 922–928.

[4] International energy agency, 2021, https://www.iea.org/reports/renewables-2021/executive-summary (Accessed 8 February 2023).

[5] International energy agency, 2021, https://www.iea.org/reports/solar-pv (Accessed 8 February 2023).

[6] The government of the Hong Kong (SAR) press releases, 2022, https://www.info.gov.hk/gia/general/202204/26/P2022042600448.htm (Accessed 8 February 2023).

[7] A. Walch, R. Castello, N. Mohajeri, JL. Scartezzini, A fast machine learning model for large-scale estimation of annual solar irradiation on rooftops, in: ISES Solar World Congress, Vpl. 45, 2020, p. 1:10.

[8] N. Mohajeri, D. Assouline, B. Guiboud, A. Bill, A. Gudmundsson, JL. Scartezzini, A city-scale roof shape classification using machine learning for solar energy applications, Renew. Energy 121 (2018) 81–93.

[9] J. Sarralde, D. Quinn, D. Wiesmann, K. Steemers, Solar energy and urban morphology: Scenarios for increasing the renewable energy potential of neighbourhoods in London, Renew. energy 73 (2015) 10–17.

[10] V. Cheng, K. Steemers, M. Montavon, R. Compagnon, Urban form, density and solar potential, in: The 23rd Conference on Passive and Low Energy Architecture 2006, 2006, pp. 1–6.

[11] C. Chatzipoulka, R. Compagnon, M. Nikolopoulou, Urban geometry and solar availability on façades and ground of real urban forms: Using London as a case study, Sol. Energy 138 (2016) 53–66.

[12] D. Robinson, Urban morphology and indicators of radiation availability, Sol. Energy 80 (2006) 1643–1648.

[13] T. Martins, L. Adolphe, L. Bastos, From solar constraints to urban design opportunities: Optimization of built form typologies in a Brazilian tropical city, Energy Build. 76 (2014) 43–56.

[14] N. Mohajeri, G. Upadhyay, A. Gudmundsson, D. Assouline, J. Kämpf, J. Scartezzini, Effects of urban compactness on solar energy potential, Renew. Energy 93 (2016) 469–482.
[15] R. Zhu, SW. Man, L. You, Santi, C. Ratti, The effect of urban morphology on the solar capacity of three-dimensional cities, Renew. Energy 153 (2020) 1111–1126.
[16] KH. Poon, JH. Kämpf, SER. Tay, NH. Wong, TG. Reindl, Parametric study of URBAN morphology on building solar energy potential in Singapore context, Urban Clim. 33 (2020) 100624.
[17] L. Ko, JC. Wang, CY. Chen, HY. Tsai, Evaluation of the development potential of rooftop solar photovoltaic in Taiwan, Renew. Energy 76 (2015) 582–595.
[18] P. Rich, R. Dubayah, W. Hetrick, S. Saving, Using viewshed models to calculate intercepted solar radiation: Applications in ecology, Am. Soc. Photogramm. Rem. Sens. Tech. Pap. Am. Soc. Photogramm. Rem. Sens. 1 (1994) 524–529.
[19] M. Morganti, A. Salvati, H. Coch, C. Cecere, Urban morphology indicators for solar energy analysis, Energy Procedia 134 (2017) 807–814.
[20] J. Yang, Y. Yang, D. Sun, C. Jin, X. Xiao, Influence of urban morphological characteristics on thermal environment, Sustain. Cities Soc. 72 (2021) 103045.
[21] S. Chen, NH. 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.
[22] H. Leng, X. Chen, Y. Ma, N. Wong, T. Ming, Urban morphology and building heating energy consumption: Evidence from Harbin, a severe cold region city, Energy Build. 224 (2020) 110143.
[23] K. Javanroodi, M. Mahdavinejad, VM. Nik, Impacts of urban morphology on reducing cooling load and increasing ventilation potential in hot-arid climate, Appl. Energy 231 (2018) 714–746.
[24] I. Lima, V. Scalco, R. Lamberts, Estimating the impact of urban densification on high-rise office building cooling loads in a hot and humid climate, Energy Build. 182 (2018) 30–44.
[25] R. Wei, D. Song, NH. Wong, M. Martin, Impact of urban morphology parameters on microclimate, Procedia Eng. 169 (2016) 142–149.
[26] X. Yong, R. Chao, P. Ma, J. Ho, Ng E., Urban morphology detection and computation for urban climate research, Landscape Urban Plan. 167 (2017) 212–224.
[27] M. Fleischmann, A. Feliciotti, O. Romice, S. Porta, Morphological Tessellation as a way of partitioning space: Improving consistency in urban morphology at the plot scale, Comput. Environ. Urban Syst. 80 (2020) 101441.
[28] A. Boccalatte, M. Thebault, C. Ménézo, J. Ramousse, M. Fossa, Evaluating the impact of urban morphology on rooftop solar radiation: A new city-scale approach based on Geneva GIS data, Energy Build. 260 (2022) 111919.
[29] AAA. Gassar, SH. Cha, Review of geographic information systems-based rooftop solar photovoltaic potential estimation approaches at urban scales, Appl. Energy 291 (2021) 116817.
[30] S. Izquierdo, C. Montañés, C. Dopazo, N. Fueyo, Roof-top solar energy potential under performance-based building energy codes: The case of Spain, Sol. Energy 85 (2011) 208–213.
[31] S. Izquierdo, M. Rodrigues, Fueyo N., A method for estimating the geographical distribution of the available roof surface area for large-scale photovoltaic energy-potential evaluations, Sol. Energy 82 (2008) 929–939.
[32] H. Fathizad, MH. Mobin, A. Gholamnia, H. Sodaiezadeh, Modeling and mapping of solar radiation using geostatistical analysis methods in Iran, Arab. J. Geosci. 10 (2017) 1–13.
[33] T. Mishra, A. Rabha, U. Kumar, K. Arunachalam, V. Sridhar, Assessment of solar power potential in a hill state of India using remote sensing and geographic information system - ScienceDirect, Rem. Sens. Appl. Soc. Environ. 19 (2020) 100370.
[34] H. Saadaoui, A. Ghennioui, B. Ikken, H. Rhinane, M. Maanan, Using GIS and photogrammetry for assessing solar photovoltaic potential on flat roofs in urban area case of the city of Ben Guerir/Morocco, Int. Arch. Photogramm. Rem. Sens. Spatial Inform. Sci. 42 (2019) 155–166.
[35] T. Hong, M. Lee, C. Koo, K. Jeong, J. Kim, Development of a method for estimating the rooftop solar photovoltaic (PV) potential by analyzing the available rooftop area using Hillshade analysis, Appl. Energy 194 (2017) 320–332.
[36] S. Tabik, A. Villegas, EL. Zapata, LF. Romero, A fast GIS-tool to compute the maximum solar energy on very large terrains, Procedia Comput. Sci. 9 (2012) 364–372.
[37] X. Liao, R. Zhu, MS. Wong, Simplified estimation modeling of land surface solar irradiation: A comparative study in Australia and China, Sustain. Energy Technol. Assess. 52 (2022) 102323.
[38] D. Assouline, N. Mohajeri, JL. Scartezzini, Large-scale rooftop solar photovoltaic technical potential estimation using random forests, Appl. Energy 217 (2018) 189–211.
[39] J. Wang, P. Li, R. Ran, Y. Che, Y. Zhou, A short-term photovoltaic power prediction model based on the gradient boost decision tree, Appl. Sci. 8 (2018) 689.
[40] SM. Babbar, CY. Lau, KF. Thang, Long term solar power generation prediction using adaboost as a hybrid of linear and non-linear machine learning model, Int. J. Adv. Comput. Sci. Appl. 12 (2021) 11.
[41] NASA power, 2021, https://power.larc.nasa.gov/data-access-viewer/ (Accessed 8 February 2023).
[42] Hong Kong observatory, 2023, https://www.hko.gov.hk/en/cis/climat.htm (Accessed 8 February 2023).
[43] H. Tong, A. Walton, J. Sang, JC. Chan, Numerical simulation of the urban boundary layer over the complex terrain of Hong Kong, Atmosp. environ. 39 (2005) 3549–3563.
[44] R. Buffat, S. Grassi, M. Raubal, A scalable method for estimating rooftop solar irradiation potential over large regions, Appl. Energy 216 (2018) 389–401.
[45] JR. Nelson, TH. Grubesic, The use of LiDAR versus unmanned aerial systems (UAS) to assess rooftop solar energy potential, Sustain. Cities Soc. 61 (2020) 102353.
[46] Hong kong building rooftop solar map, 2022, https://solarmap.emsd.gov.hk/map (Accessed 8 February 2023).
[47] MS. Wong, R. Zhu, Z. Liu, L. Lu, J. Peng, Z. Tang, CH. Lo, WK. Chan, Estimation of Hong Kongś solar energy potential using GIS and remote sensing technologies, Renew. Energy 99 (2016) 325–335.
[48] D. Li, G. Liu, S. Liao, Solar potential in urban residential buildings, Sol. Energy 111 (2015) 225–235.
[49] R. Zhu, C. Cheng, P. Santi, M. Chen, X. Zhang, M. Mazzarello, M. Wong, C. Ratti, Optimization of photovoltaic provision in a three-dimensional city using real-time electricity demand, Appl. Energy (2022).
[50] R. Zhu, L. You, P. Santi, S. Man, C. Ratti, Solar accessibility in developing cities: A case study in Kowloon east, Hong Kong, Sustain. Cities Soc. 51 (2019) 101738.
[51] M. Tanu, W. Amponsah, B. Yahaya, E. Bessah, SO. Ansah, CS. Wemegah, WA. Agyare, Evaluation of global solar radiation, cloudiness index and sky view factor as potential indicators of ghana's solar energy resource, Scient. Afr. 14 (2021) e01061.
[52] C. López, M. Sala, L. Tagliabue, F. Frontini, S. Bouziri, Solar radiation and daylighting assessment using the sky-view factor (SVF) analysis as method to evaluate urban planning densification policies impacts, Energy Procedia 91 (2016) 989–996.
[53] C. Chatzipoulka, R. Compagnon, J. Kaempf, M. Nikolopoulou, Sky view factor as predictor of solar availability on building façades, Sol. Energy 170 (2018) 1026–1038.
[54] M. Fleischmann, MOMEPY: Urban morphology measuring toolkit, J. Open Source Softw. 4 (2019) 1807.
[55] Greenhall, geopandas/geopandas: v0.6.1; 2019, 2023, http://dx.doi.org/10.5281/ZENODO.3483425 (Accessed 8 February 2023).
[56] SJ. Rey, L. Anselin, PySAL: A python library of spatial analytical methods, in: Handbook of Applied Spatial Analysis, vol. 1, Springer, 2010, pp. 175–193.
[57] A. Hagberg, PS. Swart, D. Chult, Exploring Network Structure, Dynamics, and Function using NetworkX, Vol. 8, Springer, 2008, p. 5495.
[58] A. Edussuriya, B. Chan, C. Ye, Urban morphology and air quality in dense residential environments in Hong Kong. Part I: District-level analysis, Atmos. Environ. 45 (2011) 4789–4803.
[59] E. Ng, Y. Chao, C. Liang, R. Chao, J. Fung, Improving the wind environment in high-density cities by understanding urban morphology and surface roughness: A study in Hong Kong, Landscape Urban Plan. 101 (2011) 59–74.
[60] MR. Segal, Machine learning benchmarks and random forest regression, Center Bioinform. Mol. Biostat. 1 (2004) 1–14.
[61] B. Efron, R. Tibshirani, The bootstrap method for assessing statistical accuracy, Behaviormetrika 12 (1985) 1–35.
[62] JH. Friedman, Greedy function approximation: A gradient boosting machine, Ann. statist. 1 (2001) 1189–1232.
[63] R. Ďuračiová, F. Pružinec, Effects of terrain parameters and spatial resolution of a digital elevation model on the calculation of potential solar radiation in the mountain environment: A case study of the Tatra mountains, ISPRS Int. J. Geo-Inf. 11 (2022) 389.
[64] RE. Schapire, Y. Singer, BoosTexter: A boosting-based system for text categorization, Mach. Learn. 39 (2000) 135–168.
[65] P. Kundur, J. Paserba, V. Ajjarapu, G. Andersson, A. Bose, C. Canizares, N. Hatziargyriou, D. Hill, A. Stankovic, C. Taylor, Definition and classification of power system stability IEEE/CIGRE joint task force on stability terms and definitions, IEEE Trans. Power Syst. 19 (2004) 1387–1401.
[66] NG. Baltas, P. Mazidi, J. Ma, F. de Asis Fernandez, P. Rodriguez, A comparative analysis of decision trees, support vector machines and artificial neural networks for on-line transient stability assessment, in: 2018 International Conference on Smart Energy Systems and Technologies,Vol. 1, SEST, 2018, pp. 1–6.
[67] J. Neter, W. Wasserman, MH. Kutner, Applied linear statistical models, Technometrics 39 (1996) 880.

[68] C. Dewi, RC. Chen, Random forest and support vector machine on features selection for regression analysis, Int. J. Innov. Comput. Inform. Control 15 (2019) 2027–2037.
[69] K. Bhanujyothi, K. Himabindu, D. Suryanarayana, A comparative study of random forest & K - nearest neighbors on HAR dataset using caret, Int. J. Innov. Res. Technol. 3 (2014) 6–9.
[70] U. Grömping, Variable importance assessment in regression: Linear regression versus random forest, Amer. Statist. 63 (2009) 308–319.
[71] J. Rodríguez, A. Pérez, J. Lozano, Sensitivity analysis of k-fold cross validation in prediction error estimation, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2009) 569–575.
[72] M. Feurer, F. Hutter, Hyperparameter optimization, in: Automated machine learning: Methods, systems, challenges, Vol. 1, 2019, pp. 3–33.
[73] A. Gastli, Y. Charabi, Solar electricity prospects in Oman using GIS-based solar radiation maps, Renew. Sustain. Energy Rev. 14 (2010) 790–797.